

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
23 May 2002 (23.05.2002)

PCT

(10) International Publication Number  
WO 02/40671 A2

(51) International Patent Classification<sup>7</sup>: C12N 15/12,  
C07K 14/705, 16/28, A01K 67/027, A61K 38/16, C12N  
5/10, G01N 33/50, 33/53

CA 94404 (US). GURURAJAN, Rajagopal [IN/US];  
5591 Dent Avenue, San Jose, CA 95118 (US). GANDHI,  
Ameena, R. [US/US]; 705 5th Avenue, San Francisco, CA  
94118 (US). ARVIZU, Chandra [US/US]; 490 Sherwood  
Way #1, Menlo Park, CA 94025 (US). YAO, Monique, G.  
[US/US]; 1189 Woodgate Drive, Carmel, IN 46033 (US).

(21) International Application Number: PCT/US01/44974

(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics,  
Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(22) International Filing Date:  
15 November 2001 (15.11.2001)

(81) Designated States (national): AE, AG, AL, AM, AT, AU,  
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,  
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,  
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,  
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,  
MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK,  
SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA,  
ZW.

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/249,645 16 November 2000 (16.11.2000) US

(84) Designated States (regional): ARIPO patent (GH, GM,  
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),  
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),  
European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR,  
GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent  
(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR,  
NE, SN, TD, TG).

(71) Applicant (for all designated States except US): INCYTE  
GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo  
Alto, CA 94304 (US).

Published:

— without international search report and to be republished  
upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(72) Inventors; and  
(75) Inventors/Applicants (for US only): BAUGHN, Mariah,  
R. [US/US]; 14244 Santiago Road, San Leandro, CA  
94577 (US). LU, Dyung, Aina, M. [US/US]; 233 Coy  
Drive, San Jose, CA 95123 (US). YUE, Henry [US/US];  
826 Lois Avenue, Sunnyvale, CA 94087 (US). ELLIOTT,  
Vicki, S. [US/US]; 3770 Polton Place Way, San Jose, CA  
95121 (US). THANGAVELU, Kavitha [IN/US]; 1950  
Montecito Avenue, #23, Mountain View, CA 94043 (US).  
RAMKUMAR, Jayalaxmi [IN/US]; 34359 Maybird  
Circle, Fremont, CA 94555 (US). LU, Yan [CN/US]; 3885  
Corrina Way, Palo Alto, CA 94303 (US). LO, Terrence, P.  
[CA/US]; 1451 Beach Park Boulevard, #115, Foster City,



WO 02/40671 A2

(54) Title: IMMUNOGLOBULIN SUPERFAMILY PROTEINS

(57) Abstract: The invention provides human immunoglobulin superfamily proteins (IGSFP) and polynucleotides which identify and encode IGSFP. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of IGSFP.

## IMMUNOGLOBULIN SUPERFAMILY PROTEINS

## TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of immunoglobulin superfamily proteins and to the use of these sequences in the diagnosis, treatment, and prevention of immune system disorders, neurological disorders, developmental disorders, muscle disorders, and cell proliferative disorders, including cancer, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of immunoglobulin superfamily proteins.

10

## BACKGROUND OF THE INVENTION

Most cell surface and soluble molecules that mediate functions such as recognition, adhesion or binding have evolved from a common evolutionary precursor (i.e., these proteins have structural homology). A number of molecules outside the immune system that have similar functions are also derived from this same evolutionary precursor. These molecules are classified as members of the immunoglobulin (Ig) superfamily. The criteria for a protein to be a member of the Ig superfamily is to have one or more Ig domains, which are regions of 70-110 amino acid residues in length homologous to either Ig variable-like (V) or Ig constant-like (C) domains. Members of the Ig superfamily include antibodies (Ab), T cell receptors (TCRs), class I and II major histocompatibility (MHC) proteins, CD2, CD3, CD4, CD8, poly-Ig receptors, Fc receptors, neural cell-adhesion molecule (NCAM) and platelet-derived growth factor receptor (PDGFR).

Ig domains (V and C) are regions of conserved amino acid residues that give a polypeptide a globular tertiary structure called an immunoglobulin (or antibody) fold, which consists of two approximately parallel layers of  $\beta$ -sheets. Conserved cysteine residues form an intrachain disulfide-bonded loop, 55-75 amino acid residues in length, which connects the two layers of the  $\beta$ -sheets.

Each  $\beta$ -sheet has three or four anti-parallel  $\beta$ -strands of 5-10 amino acid residues. Hydrophobic and hydrophilic interactions of amino acid residues within the  $\beta$ -strands stabilize the Ig fold (hydrophobic on inward facing amino acid residues and hydrophilic on the amino acid residues in the outward facing portion of the strands). A V domain consists of a longer polypeptide than a C domain, with an additional pair of  $\beta$ -strands in the Ig fold.

A consistent feature of Ig superfamily genes is that each sequence of an Ig domain is encoded by a single exon. It is possible that the superfamily evolved from a gene coding for a single Ig domain involved in mediating cell-cell interactions. New members of the superfamily then arose by exon and gene duplications. Modern Ig superfamily proteins contain different numbers of V and/or C

domains. Another evolutionary feature of this superfamily is the ability to undergo DNA rearrangements, a unique feature retained by the antigen receptor members of the family.

Many members of the Ig superfamily are integral plasma membrane proteins with extracellular Ig domains. The hydrophobic amino acid residues of their transmembrane domains and 5 their cytoplasmic tails are very diverse, with little or no homology among Ig family members or to known signal-transducing structures. There are exceptions to this general superfamily description. For example, the cytoplasmic tail of PDGFR has tyrosine kinase activity. In addition Thy-1 is a glycoprotein found on thymocytes and T cells. This protein has no cytoplasmic tail, but is instead attached to the plasma membrane by a covalent glycoprophosphatidylinositol linkage.

10 Another common feature of many Ig superfamily proteins is the interactions between Ig domains which are essential for the function of these molecules. Interactions between Ig domains of a multimeric protein can be either homophilic or heterophilic (i.e., between the same or different Ig domains). Antibodies are multimeric proteins which have both homophilic and heterophilic interactions between Ig domains. Pairing of constant regions of heavy chains forms the Fc region of an antibody 15 and pairing of variable regions of light and heavy chains form the antigen binding site of an antibody. Heterophilic interactions also occur between Ig domains of different molecules. These interactions provide adhesion between cells for significant cell-cell interactions in the immune system and in the developing and mature nervous system. (Reviewed in Abbas, A.K. et al. (1991) Cellular and Molecular Immunology, W.B. Saunders Company, Philadelphia PA, pp.142-145.)

20 Antibodies

Antibodies are multimeric members of the Ig superfamily which are either expressed on the surface of B-cells or secreted by B-cells into the circulation. Antibodies bind and neutralize foreign 25 antigens in the blood and other extracellular fluids. The prototypical antibody is a tetramer consisting of two identical heavy polypeptide chains (H-chains) and two identical light polypeptide chains (L-chains) interlinked by disulfide bonds. This arrangement confers the characteristic Y-shape to antibody molecules. Antibodies are classified based on their H-chain composition. The five antibody classes, IgA, IgD, IgE, IgG and IgM, are defined by the  $\alpha$ ,  $\delta$ ,  $\epsilon$ ,  $\gamma$ , and  $\mu$  H-chain types. There are two types of L-chains,  $\kappa$  and  $\lambda$ , either of which may associate as a pair with any H-chain pair. IgG, the most common class of antibody found in the circulation, is tetrameric, while the other classes of 30 antibodies are generally variants or multimers of this basic structure.

H-chains and L-chains each contain an N-terminal variable region and a C-terminal constant region. The constant region consists of about 110 amino acids in L-chains and about 330 or 440 amino acids in H-chains. The amino acid sequence of the constant region is nearly identical among H- or L-

chains of a particular class. The variable region consists of about 110 amino acids in both H- and L-chains. However, the amino acid sequence of the variable region differs among H- or L-chains of a particular class. Within each H- or L-chain variable region are three hypervariable regions of extensive sequence diversity, each consisting of about 5 to 10 amino acids. In the antibody molecule, 5 the H- and L-chain hypervariable regions come together to form the antigen recognition site.

(Reviewed in Alberts, B. et al. (1994) Molecular Biology of the Cell, Garland Publishing, New York NY, pp. 1206-1213 and 1216-1217.)

Both H-chains and L-chains contain the repeated Ig domains of members of the Ig superfamily. For example, a typical H-chain contains four Ig domains, three of which occur within the 10 constant region and one of which occurs within the variable region and contributes to the formation of the antigen recognition site. Likewise, a typical L-chain contains two Ig domains, one of which occurs within the constant region and one of which occurs within the variable region.

The immune system is capable of recognizing and responding to any foreign molecule that enters the body. Therefore, the immune system must be armed with a full repertoire of antibodies 15 against all potential antigens. Such antibody diversity is generated by somatic rearrangement of gene segments encoding variable and constant regions. These gene segments are joined together by site-specific recombination which occurs between highly conserved DNA sequences that flank each gene segment. Because there are hundreds of different gene segments, millions of unique genes can be generated combinatorially. In addition, imprecise joining of these segments and an unusually high rate 20 of somatic mutation within these segments further contribute to the generation of a diverse antibody population.

#### Neural Cell Adhesion Proteins

Neural cell adhesion proteins (NCAPs) play roles in the establishment of neural networks during development and regeneration of the nervous system (Uyemura et al. (1996) Essays Biochem. 25 31:37-48; Brummendorf and Rathjen (1996) Curr. Opin. Neurobiol. 6:584-593). NCAP participates in neuronal cell migration, cell adhesion, neurite outgrowth, axonal fasciculation, pathfinding, synaptic target-recognition, synaptic formation, myelination and regeneration. NCAPs are expressed on the surfaces of neurons associated with learning and memory. Mutations in genes encoding NCAPs are linked with neurological diseases, including Charcot-Marie-Tooth disease (a hereditary neuropathy), 30 Dejerine-Sottas disease, X-linked hydrocephalus, MASA syndrome (mental retardation, aphasia, shuffling gait and adducted thumbs), and spastic paraplegia type I. In some cases, expression of NCAP is not restricted to the nervous system. L1, for example, is expressed in melanoma cells and hematopoietic tumor cells where it is implicated in cell spreading and migration, and may play a role in

tumor progression (Montgomery et al. (1996) *J. Cell Biol.* 132:475-485).

NCAPs have at least one immunoglobulin constant or variable domain (Uyemura et al., *supra*). They are generally linked to the plasma membrane through a transmembrane domain and/or a glycosyl-phosphatidylinositol (GPI) anchor. The GPI linkage can be cleaved by GPI phospholipase C.

5 Most NCAPs consist of an extracellular region made up of one or more immunoglobulin domains, a membrane spanning domain, and an intracellular region. Many NCAPs contain post-translational modifications including covalently attached oligosaccharide, glucuronic acid, and sulfate. NCAPs fall into three subgroups: simple-type, complex-type, and mixed-type. Simple-type NCAPs contain one or more variable or constant immunoglobulin domains, but lack other types of domains. Members of the

10 simple-type subgroup include Schwann cell myelin protein (SMP), limbic system-associated membrane protein (LAMP) and opiate-binding cell-adhesion molecule (OBCAM). The complex-type NCAPs contain fibronectin type III domains in addition to the immunoglobulin domains. The complex-type subgroup includes neural cell-adhesion molecule (NCAM), axonin-1, F11, Bravo, and L1. Mixed-type NCAPs contain a combination of immunoglobulin domains and other motifs such as tyrosine kinase,

15 epidermal growth factor-like, sema, and PSI (plexins, semaphorins, and integrins) domains. This subgroup includes Trk receptors of nerve growth factors such as nerve growth factor (NGF) and neurotropin 4 (NT4), Neu differentiation factors such as glial growth factor II (GGFII) and acetylcholine receptor-inducing factor (ARIA), the semaphorin/collapsin family such as semaphorin B and collapsin, and receptors for members of the semaphorin/collapsin family such as plexin (for plexin,

20 see below).

An NCAP subfamily, the NCAP-LON subgroup, includes cell adhesion proteins expressed on distinct subpopulations of brain neurons. Members of the NCAP-LON subgroup possess three immunoglobulin domains and bind to cell membranes through GPI anchors. Kilon (a kindred of NCAP-LON), for example, is expressed in the brain cerebral cortex and hippocampus (Funatsu et al. 25 (1999) *J. Biol. Chem.* 274:8224-8230). Immunostaining localizes Kilon to the dendrites and soma of pyramidal neurons. Kilon has three C2 type immunoglobulin-like domains, six predicted glycosylation sites, and a GPI anchor. Expression of Kilon is developmentally regulated. It is expressed at higher levels in adult brain in comparison to embryonic and early postnatal brains. Confocal microscopy shows the presence of Kilon in dendrites of hypothalamic magnocellular neurons secreting

30 neuropeptides, oxytocin, or arginine vasopressin (Miyata et al. (2000) *J. Comp. Neurol.* 424:74-85). Arginine vasopressin regulates body fluid homeostasis, extracellular osmolarity and intravascular volume. Oxytocin induces contractions of uterine smooth muscle during child birth and of myoepithelial cells in mammary glands during lactation. In magnocellular neurons, Kilon is proposed to

play roles in the reorganization of dendritic connections during neuropeptide secretion.

Sidekick (SDK) is a member of the NCAP family. The extracellular region of SDK contains six immunoglobulin domains and thirteen fibronectin type III domains. SDK is involved in cell-cell interaction during eye development in *Drosophila* (Nguyen, D.N.T. et al. (1997) Development

5 124:3303).

#### Synaptic Membrane Glycoproteins

Specialized cell junctions can occur at points of cell-cell contact. Among these cell junctions are communicating junctions which mediate the passage of chemical and electrical signals between cells. In the central nervous system, communicating junctions between neurons are known as synaptic junctions. They are composed of the membranes and cytoskeletons of the pre- and post-synaptic neurons. Some glycoproteins, found in biochemically isolated synaptic subfractions such as the synaptic membrane (SM) and postsynaptic density (PSD) fractions, have been identified and their functions established. An example is the SM glycoprotein, gp50, identified as the  $\beta$ 2 subunit of the  $\text{Na}^+/\text{K}^+$ -ATPase.

15 Two glycoproteins, gp65 and gp55, are major components of synaptic membranes prepared from rat forebrain. They are members of the Ig superfamily containing three and two Ig domains, respectively. As members of the Ig superfamily, it is proposed that a possible function of these proteins is to mediate adhesive interactions at the synaptic junction (Langnaese, K. et al. (1997) J. Biol. Chem. 272:821-827).

#### Lectins

Lectins comprise a ubiquitous family of extracellular glycoproteins which bind cell surface carbohydrates specifically and reversibly, resulting in the agglutination of cells (reviewed in Drickamer, K. and Taylor, M.E. (1993) Annu. Rev. Cell Biol. 9:237-264). This function is particularly important for activation of the immune response. Lectins mediate the agglutination and mitogenic 25 stimulation of lymphocytes at sites of inflammation (Lasky, L.A. (1991) J. Cell. Biochem. 45:139-146; Paietta, E. et al. (1989) J. Immunol. 143:2850-2857).

30 Sialic acid binding Ig-like lectins (SIGLECs) are members of the Ig superfamily that bind to sialic acids in glycoproteins and glycolipids. SIGLECs include sialoadhesin, CD22, CD33, myelin-associated glycoprotein (MAG), SIGLEC-5, SIGLEC-6, SIGLEC-7, and SIGLEC-8. The extracellular region of SIGLEC has a membrane distal V-set domain followed by varying numbers of C2-set domains. The sialic acid binding domain is mapped to the V-set domain. Except for MAG which is expressed exclusively in the nervous system, most SIGLECs are expressed on distinct subsets of hemopoietic cells. For example, SIGLEC-8 is expressed exclusively in eosinophils, one form of

polymorphonuclear leucocyte (granulocyte) (Floyd, H. et al. (2000) *J. Biol. Chem.* 275: 861-866).

#### Leucine-Rich Repeat Proteins

Leucine-rich repeat proteins (LRRPs) are involved in protein-protein interactions. LRRPs such as mammalian neuronal leucine-rich repeat proteins (NLLR-1 and NLLR-2), *Drosophila* connectin, slit, chaopin, and toll all play roles in neuronal development. The extracellular region of LRRPs contains varying numbers of leucine-rich repeats, immunoglobulin-like domains, and fibronectin type III domains (Taguchi, A. et al. (1996) *Brain Res. Mol. Brain Res.* 35:31-40).

In addition to the V and C2 sets of immunoglobulin-like domains, there is a D set immunoglobulin-like domain, named IPT/TIG (for immunoglobulin-like fold shared by plexins and transcription factors). IPT/TIG containing proteins include plexins, MET/RON/SEA (hepatocyte growth factor receptor family), and the transcription factor XCoe2, a transcription factor of the Col/Olf-1/EBF family involved in the specification of primary neurons in *Xenopus* (Bork, P. et al. (1999) *Trends Biochem. Sci.* 24:261-263; Santoro, N.M. et al. (1996) *Mol. Cell Biol.* 16:7072-7083; Dubois L. et al. (1998) *Curr. Biol.* 8:199-209). Plexins such as plexin A and VESPR have been shown to be neuronal semaphorin receptors that control axon guidance (Winberg M.L. et al. (1998) *Cell* 95:903-916).

The discovery of new immunoglobulin superfamily proteins, and the polynucleotides encoding them, satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of immune system disorders, neurological disorders, developmental disorders, muscle disorders, and cell proliferative disorders, including cancer, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of immunoglobulin superfamily proteins.

#### **SUMMARY OF THE INVENTION**

The invention features purified polypeptides, immunoglobulin superfamily proteins, referred to collectively as "IGSFP" and individually as "IGSFP-1," "IGSFP-2," "IGSFP-3," "IGSFP-4," "IGSFP-5," and "IGSFP-6." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6. In one alternative, the

invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-6.

The invention further provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-6. In 10 another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:7-12.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention 20 provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, c) a 25 biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide 30 encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, b) a polypeptide comprising a naturally

occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6.

The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:7-12, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:7-12, c) 10 a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:7-12, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:7-12, c) 15 a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe 20 comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:7-12, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% 25 identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:7-12, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b)

detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and 10 a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-6. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional IGSFP, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of 20 treating a disease or condition associated with decreased expression of functional IGSFP, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group

consisting of SEQ ID NO:1-6. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of 5 treating a disease or condition associated with overexpression of functional IGSFP, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, b) a polypeptide comprising a 10 naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6. The method comprises a) combining the polypeptide with at least one 15 test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, b) a polypeptide comprising a 20 naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6. The method comprises a) combining the polypeptide with at least one 25 test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

30 The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:7-12, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting

altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 5 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:7-12, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:7-12, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of 10 ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:7-12, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide 15 sequence selected from the group consisting of SEQ ID NO:7-12, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount 20 of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

#### BRIEF DESCRIPTION OF THE TABLES

25 Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability scores for the matches between each polypeptide and its homolog(s) are also shown.

30 Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble

polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of 5 the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

### **DESCRIPTION OF THE INVENTION**

10 Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

15 It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

20 Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the 25 cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

### **DEFINITIONS**

30 "IGSFP" refers to the amino acid sequences of substantially purified IGSFP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of IGSFP. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other

compound or composition which modulates the activity of IGSFP either by directly interacting with IGSFP or by acting on components of the biological pathway in which IGSFP participates.

An "allelic variant" is an alternative form of the gene encoding IGSFP. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in 5 polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

10 "Altered" nucleic acid sequences encoding IGSFP include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as IGSFP or a polypeptide with at least one functional characteristic of IGSFP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding IGSFP, and improper or unexpected hybridization to allelic variants, with a 15 locus other than the normal chromosomal locus for the polynucleotide sequence encoding IGSFP. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent IGSFP. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological 20 or immunological activity of IGSFP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; 25 and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence 30 to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of IGSFP. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of IGSFP either by directly interacting with IGSFP or by acting on components of the biological pathway in which IGSFP 5 participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')<sub>2</sub>, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind IGSFP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used 10 to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that 15 makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

20 The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a specific molecular target. Aptamers are derived from an in vitro evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No. 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries.

Aptamer compositions may be double-stranded or single-stranded, and may include 25 deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or 2'-NH<sub>2</sub>), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system.

30 Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker. (See, e.g., Brody, E.N. and L. Gold (2000) J. Biotechnol. 74:5-13.)

The term "intramer" refers to an aptamer which is expressed in vivo. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels

in the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl Acad. Sci. USA 96:3606-3610).

The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on 5 substrates containing right-handed nucleotides.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified 10 sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring 15 nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic IGSFP, or of any oligopeptide thereof, 20 to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

25 A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding IGSFP or fragments of IGSFP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be 30 associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to

repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

5 "Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as 10 conservative amino acid substitutions.

	Original Residue	Conservative Substitution
15	Ala	Gly, Ser
	Arg	His, Lys
	Asn	Asp, Gln, His
	Asp	Asn, Glu
	Cys	Ala, Ser
	Gln	Asn, Glu, His
	Glu	Asp, Gln, His
20	Gly	Ala
	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val
	Leu	Ile, Val
	Lys	Arg, Gln, Glu
	Met	Leu, Ile
	Phe	His, Met, Leu, Trp, Tyr
25	Ser	Cys, Thr
	Thr	Ser, Val
	Trp	Phe, Tyr
	Tyr	His, Phe, Trp
	Val	Ile, Leu, Thr

30 Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

35 A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide.

40 Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an

alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

5 A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

10 "Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

"Exon shuffling" refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

15 A "fragment" is a unique portion of IGSFP or the polynucleotide encoding IGSFP which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the 25 specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

30 A fragment of SEQ ID NO:7-12 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:7-12, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:7-12 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:7-12 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:7-12 and the region of SEQ ID NO:7-12 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-6 is encoded by a fragment of SEQ ID NO:7-12. A fragment of SEQ ID NO:1-6 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-6. For example, a fragment of SEQ ID NO:1-6 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-6. The precise length of a 5 fragment of SEQ ID NO:1-6 and the region of SEQ ID NO:1-6 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full 10 length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a 15 standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e 20 sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polymucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue 25 weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment 30 Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at <http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2

Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to 5 compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

*Matrix: BLOSUM62*

*Reward for match: 1*

*Penalty for mismatch: -2*

10 *Open Gap: 5 and Extension Gap: 2 penalties*

*Gap x drop-off: 50*

*Expect: 10*

*Word Size: 11*

*Filter: on*

15 Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported 20 by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

25 Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

30 The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e

sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and “diagonals saved”=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by 5 CLUSTAL V as the “percent similarity” between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the “BLAST 2 Sequences” tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

10 *Matrix: BLOSUM62*

*Open Gap: 11 and Extension Gap: 1 penalties*

*Gap x drop-off: 50*

*Expect: 10*

*Word Size: 3*

15 *Filter: on*

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 20 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

25 “Human artificial chromosomes” (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term “humanized antibody” refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

30 “Hybridization” refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the “washing” step(s). The washing step(s) is particularly important in determining the

stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be 5 varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 10 5°C to 20°C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength and pH. The  $T_m$  is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating  $T_m$  and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; 15 specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking 20 reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency 25 conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g.,  $C_0t$  or  $R_0t$  analysis) or formed between one 30 nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide

sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect 5 cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of IGSFP which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of IGSFP which is useful in any of the antibody production methods disclosed herein or known in the art.

10 The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

15 The term "modulate" refers to a change in the activity of IGSFP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of IGSFP.

20 The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

25 "Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

30 "Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an IGSFP may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary

by cell type depending on the enzymatic milieu of IGSFP.

"Probe" refers to nucleic acid sequences encoding IGSFP, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule.

5 Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

10 Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the 15 specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR 20. Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such 25 purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU 30 primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to

avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, 10 as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. 15 This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, *supra*. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. 20 Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be used to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated 25 regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, 30 chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of

the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing IGSFP, nucleic acids encoding IGSFP, or fragments thereof may comprise a bodily fluid; an extract from a 5 cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure 10 of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are 15 removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

20 "Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" or "expression profile" refers to the collective pattern of gene expression 25 by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based 30 on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently

transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The 5 nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, 10 fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having 15 at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater 20 sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the 25 reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The 30 presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of

the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

## THE INVENTION

The invention is based on the discovery of new human immunoglobulin superfamily proteins (IGSFP), the polynucleotides encoding IGSFP, and the use of these compositions for the diagnosis, treatment, or prevention of immune system disorders, neurological disorders, developmental disorders, muscle disorders, and cell proliferative disorders, including cancer.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (GenBank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability scores for the matches between each polypeptide and its homolog(s). Column 5 shows the annotation of the GenBank homolog(s) along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7

shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are immunoglobulin superfamily proteins. For 5 example, SEQ ID NO:1 is 87% identical to murine punc (putative neural cell adhesion protein) (GenBank ID g3068592) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:1 also contains immunoglobulin and fibronectin type III domains as determined by searching for statistically significant matches in the 10 hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS analysis provides further corroborative evidence that SEQ ID NO:1 is a neural cell adhesion protein. SEQ ID NO:2-6 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-6 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were 15 assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and the corresponding Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention. Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments of 20 the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:7-12 or that distinguish between SEQ ID NO:7-12 and related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length 25 polynucleotide sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences in column 5 relative to their respective full length sequences.

The identification numbers in Column 5 of Table 4 may refer specifically, for example, to 30 Incyte cDNAs along with their corresponding cDNA libraries. For example, 7184821H1 is the identification number of an Incyte cDNA sequence, and BONRFEC01 is the cDNA library from which it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries (e.g., 71553470V1). Alternatively, the identification numbers in column 5 may refer to GenBank cDNAs or ESTs (e.g., g2993996) which contributed to the assembly of the full

length polynucleotide sequences. In addition, the identification numbers in column 5 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation “ENST”). Alternatively, the identification numbers in column 5 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation “NM” or “NT”) or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation “NP”). Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an “exon stitching” algorithm. For example, *FL\_XXXXXX\_N<sub>1</sub>\_N<sub>2</sub>\_YYYYY\_N<sub>3</sub>\_N<sub>4</sub>* represents a “stitched” sequence in which *XXXXXX* is the identification number of the cluster of sequences to which the algorithm was applied, and *YYYYY* is the number of the prediction generated by the algorithm, and *N<sub>1,2,3..</sub>*, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the identification numbers in column 5 may refer to assemblages of exons brought together by an “exon-stretching” algorithm. For example, *FLXXXXXX\_gAAAAAA\_gBBBBBB\_1\_N* is the identification number of a “stretched” sequence, with *XXXXXX* being the Incyte project identification number, *gAAAAAA* being the GenBank identification number of the human genomic sequence to which the “exon-stretching” algorithm was applied, *gBBBBBB* being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and *N* referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the “exon-stretching” algorithm, a RefSeq identifier (denoted by “NM,” “NP,” or “NT”) may be used in place of the GenBank identifier (*i.e.*, *gBBBBBB*).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK)
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).

INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.
------	---

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in column 5 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte 5 cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and 10 vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses IGSFP variants. A preferred IGSFP variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the IGSFP amino acid sequence, and which contains at least one functional or structural characteristic of IGSFP.

15 The invention also encompasses polynucleotides which encode IGSFP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:7-12, which encodes IGSFP. The polynucleotide sequences of SEQ ID NO:7-12, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar 20 backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding IGSFP. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding IGSFP. A particular aspect of the invention encompasses a variant of a 25 polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:7-12 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:7-12. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of IGSFP.

30 In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant of a polynucleotide sequence encoding IGSFP. A splice variant may have portions which have significant

sequence identity to the polynucleotide sequence encoding IGSFP, but will generally have a greater or lesser number of polynucleotides due to additions or deletions of blocks of sequence arising from alternate splicing of exons during mRNA processing. A splice variant may have less than about 70%, or alternatively less than about 60%, or alternatively less than about 50% polynucleotide sequence

5 identity to the polynucleotide sequence encoding IGSFP over its entire length; however, portions of the splice variant will have at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or alternatively 100% polynucleotide sequence identity to portions of the polynucleotide sequence encoding IGSFP. Any one of the splice variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of IGSFP.

10 It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding IGSFP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These

15 combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring IGSFP, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode IGSFP and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring IGSFP under appropriately selected

20 conditions of stringency, it may be advantageous to produce nucleotide sequences encoding IGSFP or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide

25 sequence encoding IGSFP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode IGSFP and IGSFP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the

30 synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding IGSFP or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of

hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:7-12 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

5 Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE 10 amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system 15 (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding IGSFP may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) *PCR Methods Applic.* 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown 25 sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) *Nucleic Acids Res.* 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) *PCR Methods Applic.* 1:111-119.) In this method, multiple restriction enzyme digestions and 30 ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) *Nucleic Acids Res.* 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo

Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, 5 to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) 10 library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments 20 which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode IGSFP may be cloned in recombinant DNA molecules that direct expression of IGSFP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally 25 equivalent amino acid sequence may be produced and used to express IGSFP.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter IGSFP-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic 30 oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such

as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No. 5,837,458; Chang, C.-C. et al. (1999) *Nat. Biotechnol.* 17:793-797; Christians, F.C. et al. (1999) *Nat. Biotechnol.* 17:259-264; and Crameri, A. et al. (1996) *Nat. Biotechnol.* 14:315-319) to alter or improve the biological properties of IGSFP, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding IGSFP may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) *Nucleic Acids Symp. Ser.* 7:215-223; and Horn, T. et al. (1980) *Nucleic Acids Symp. Ser.* 7:225-232.) Alternatively, IGSFP itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) *Proteins, Structures and Molecular Properties*, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) *Science* 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of IGSFP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) *Methods Enzymol.* 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, *supra*, pp. 28-53.)

In order to express a biologically active IGSFP, the nucleotide sequences encoding IGSFP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and

inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding IGSFP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding IGSFP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where 5 sequences encoding IGSFP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both 10 natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding IGSFP and appropriate transcriptional and translational control 15 elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences 20 encoding IGSFP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or 25 animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509; Engelhard, E.K. et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:3224-3227; Sandig, V. et al. (1996) *Hum. Gene Ther.* 7:1937-1945; Takamatsu, N. (1987) *EMBO J.* 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659; and 30 Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) *Cancer Gen. Ther.* 5(6):350-356; Yu, M. et al. (1993) *Proc. Natl. Acad. Sci. USA*

90(13):6340-6344; Buller, R.M. et al. (1985) *Nature* 317(6040):813-815; McGregor, D.P. et al. (1994) *Mol. Immunol.* 31(3):219-226; and Verma, I.M. and N. Somia (1997) *Nature* 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding IGSFP. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding IGSFP can be achieved using a multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding IGSFP into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for *in vitro* transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509.) When large quantities of IGSFP are needed, e.g. for the production of antibodies, vectors which direct high level expression of IGSFP may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of IGSFP. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast *Saccharomyces cerevisiae* or *Pichia pastoris*. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, *supra*; Bitter, G.A. et al. (1987) *Methods Enzymol.* 153:516-544; and Scorer, C.A. et al. (1994) *Bio/Technology* 12:181-184.)

Plant systems may also be used for expression of IGSFP. Transcription of sequences encoding IGSFP may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) *EMBO J.* 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) *EMBO J.* 3:1671-1680; Broglie, R. et al. (1984) *Science* 224:838-843; and Winter, J. et al. (1991) *Results Probl. Cell Differ.* 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., *The McGraw Hill Yearbook of Science and Technology* (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding IGSFP may be ligated into

an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses IGSFP in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, 10 or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of IGSFP in cell lines is preferred. For example, sequences encoding IGSFP can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous 15 expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue 20 culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk* and *apr* cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or 25 herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which 30 alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech),  $\beta$  glucuronidase and its substrate  $\beta$ -glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to

quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) *Methods Mol. Biol.* 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if 5 the sequence encoding IGSFP is inserted within a marker gene sequence, transformed cells containing sequences encoding IGSFP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding IGSFP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

10 In general, host cells that contain the nucleic acid sequence encoding IGSFP and that express IGSFP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

15 Immunological methods for detecting and measuring the expression of IGSFP using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on IGSFP is preferred, but a 20 competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) *Serological Methods, a Laboratory Manual*, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) *Current Protocols in Immunology*, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) *Immunochemical Protocols*, Humana Press, Totowa NJ.)

25 A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding IGSFP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding IGSFP, or any fragments thereof, may be cloned into a vector 30 for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega

(Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding IGSFP may be cultured under 5 conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode IGSFP may be designed to contain signal sequences which direct secretion of IGSFP through a prokaryotic or eukaryotic cell membrane.

10 In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells 15 which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid 20 sequences encoding IGSFP may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric IGSFP protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of IGSFP activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity 25 matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion 30 proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the IGSFP encoding sequence and the heterologous protein sequence, so that IGSFP may be cleaved away from the heterologous moiety following purification. Methods for fusion protein

expression and purification are discussed in Ausubel (1995, *supra*, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled IGSFP may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These 5 systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, <sup>35</sup>S-methionine.

IGSFP of the present invention or fragments thereof may be used to screen for compounds that specifically bind to IGSFP. At least one and up to a plurality of test compounds may be screened 10 for specific binding to IGSFP. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of IGSFP, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) *Current Protocols in Immunology* 1(2): 15 Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which IGSFP binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express IGSFP, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing IGSFP or cell membrane fractions which contain IGSFP are then contacted 20 with a test compound and binding, stimulation, or inhibition of activity of either IGSFP or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the 25 assay may comprise the steps of combining at least one test compound with IGSFP, either in solution or affixed to a solid support, and detecting the binding of IGSFP to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

30 IGSFP of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of IGSFP. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for IGSFP activity, wherein IGSFP is combined with at least one test compound, and the activity of IGSFP in the

presence of a test compound is compared with the activity of IGSFP in the absence of the test compound. A change in the activity of IGSFP in the presence of the test compound is indicative of a compound that modulates the activity of IGSFP. Alternatively, a test compound is combined with an in vitro or cell-free system comprising IGSFP under conditions suitable for IGSFP activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of IGSFP may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding IGSFP or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) *Science* 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) *Clin. Invest.* 97:1999-2002; Wagner, K.U. et al. (1997) *Nucleic Acids Res.* 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding IGSFP may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) *Science* 282:1145-1147).

Polynucleotides encoding IGSFP can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding IGSFP is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively,

a mammal inbred to overexpress IGSFP, e.g., by secreting IGSFP in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) *Biotechnol. Annu. Rev.* 4:55-74).

## **THERAPEUTICS**

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between 5 regions of IGSFP and immunoglobulin superfamily proteins. In addition, the expression of IGSFP is closely associated with eosinophils, lung tissue, prostate tumor tissue, uterine myometrial tissue, and brain tissues such as globus pallidus and substantia innominata. Therefore, IGSFP appears to play a role in immune system disorders, neurological disorders, developmental disorders, muscle disorders, and cell proliferative disorders, including cancer. In the treatment of disorders associated with 10 increased IGSFP expression or activity, it is desirable to decrease the expression or activity of IGSFP. In the treatment of disorders associated with decreased IGSFP expression or activity, it is desirable to increase the expression or activity of IGSFP.

Therefore, in one embodiment, IGSFP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or 15 activity of IGSFP. Examples of such disorders include, but are not limited to, an immune system disorder such as acquired immunodeficiency syndrome (AIDS), X-linked agammaglobulinemia of Bruton, common variable immunodeficiency (CVI), DiGeorge's syndrome (thymic hypoplasia), thymic dysplasia, isolated IgA deficiency, severe combined immunodeficiency disease (SCID), immunodeficiency with thrombocytopenia and eczema (Wiskott-Aldrich syndrome), Chediak-Higashi 20 syndrome, chronic granulomatous diseases, hereditary angioneurotic edema, immunodeficiency associated with Cushing's disease, Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, 25 dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, 30 Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a neurological disorder such as epilepsy, ischemic cerebrovascular disease,

stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural

5 empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central

10 nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD),

15 akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental

20 retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a muscle disorder such as cardiomyopathy, myocarditis, Duchenne's

25 muscular dystrophy, Becker's muscular dystrophy, myotonic dystrophy, central core disease, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, infectious myositis, polymyositis, dermatomyositis, inclusion body myositis, thyrotoxic myopathy, and ethanol myopathy; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal

30 hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis,

prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus.

In another embodiment, a vector capable of expressing IGSFP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of IGSFP including, but not limited to, those described above.

5 In a further embodiment, a composition comprising a substantially purified IGSFP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of IGSFP including, but not limited to, those provided above.

10 In still another embodiment, an agonist which modulates the activity of IGSFP may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of IGSFP including, but not limited to, those listed above.

15 In a further embodiment, an antagonist of IGSFP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of IGSFP. Examples of such disorders include, but are not limited to, those immune system disorders, neurological disorders, developmental disorders, muscle disorders, and cell proliferative disorders, including cancer, described above. In one aspect, an antibody which specifically binds IGSFP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express IGSFP.

20 In an additional embodiment, a vector expressing the complement of the polynucleotide encoding IGSFP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of IGSFP including, but not limited to, those described above.

25 In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

30 An antagonist of IGSFP may be produced using methods which are generally known in the art. In particular, purified IGSFP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind IGSFP. Antibodies to IGSFP may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments

produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with IGSFP or with any fragment or oligopeptide thereof which 5 has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

10 It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to IGSFP have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of IGSFP amino acids may be fused with those of another protein, such as KLH, and antibodies to the 15 chimeric molecule may be produced.

Monoclonal antibodies to IGSFP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) *Nature* 256:495-497; Kozbor, D. et al. (1985) *J. 20 Immunol. Methods* 81:31-42; Cote, R.J. et al. (1983) *Proc. Natl. Acad. Sci. USA* 80:2026-2030; and Cole, S.P. et al. (1984) *Mol. Cell Biol.* 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) *Proc. 25 Natl. Acad. Sci. USA* 81:6851-6855; Neuberger, M.S. et al. (1984) *Nature* 312:604-608; and Takeda, S. et al. (1985) *Nature* 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce IGSFP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, 30 D.R. (1991) *Proc. Natl. Acad. Sci. USA* 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:3833-3837; Winter,

G. et al. (1991) *Nature* 349:293-299.)

Antibody fragments which contain specific binding sites for IGSFP may also be generated. For example, such fragments include, but are not limited to,  $F(ab')_2$  fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of 5 the  $F(ab')_2$  fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) *Science* 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either 10 polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between IGSFP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering IGSFP epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

15 Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for IGSFP. Affinity is expressed as an association constant,  $K_a$ , which is defined as the molar concentration of IGSFP-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The  $K_a$  determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for 20 multiple IGSFP epitopes, represents the average affinity, or avidity, of the antibodies for IGSFP. The  $K_a$  determined for a preparation of monoclonal antibodies, which are monospecific for a particular IGSFP epitope, represents a true measure of affinity. High-affinity antibody preparations with  $K_a$  ranging from about  $10^9$  to  $10^{12}$  L/mole are preferred for use in immunoassays in which the IGSFP-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with  $K_a$  25 ranging from about  $10^6$  to  $10^7$  L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of IGSFP, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

30 The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of IGSFP-antibody

complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding IGSFP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding IGSFP. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding IGSFP. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totowa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) *J. Allergy Clin. Immunol.* 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) *Blood* 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) *Pharmacol. Ther.* 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) *Br. Med. Bull.* 51(1):217-225; Boado, R.J. et al. (1998) *J. Pharm. Sci.* 87(11):1308-1315; and Morris, M.C. et al. (1997) *Nucleic Acids Res.* 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding IGSFP may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) *Science* 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) *Science* 270:475-480; Bordignon, C. et al. (1995) *Science* 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) *Cell* 75:207-216; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:643-666; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) *Science* 270:404-410; Verma, I.M. and N. Somia (1997) *Nature* 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated

cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) *Nature* 335:395-396; Poeschla, E. et al. (1996) *Proc. Natl. Acad. Sci. USA.* 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in IGSFP expression or regulation causes disease, the expression of IGSFP from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in IGSFP are treated by constructing mammalian expression vectors encoding IGSFP and introducing these vectors by mechanical means into IGSFP-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) *Annu. Rev. Biochem.* 62:191-217; Ivics, Z. (1997) *Cell* 91:501-510; Boulay, J-L. and H. Récipon (1998) *Curr. Opin. Biotechnol.* 9:445-450).

Expression vectors that may be effective for the expression of IGSFP include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). IGSFP may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or  $\beta$ -actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) *Proc. Natl. Acad. Sci. USA* 89:5547-5551; Gossen, M. et al. (1995) *Science* 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) *Curr. Opin. Biotechnol.* 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, supra), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding IGSFP from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polymucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method

(Graham, F.L. and A.J. Eb (1973) *Virology* 52:456-467), or by electroporation (Neumann, E. et al. (1982) *EMBO J.* 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to IGSFP expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding IGSFP under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) *Proc. Natl. Acad. Sci. USA* 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) *J. Virol.* 61:1647-1650; Bender, M.A. et al. (1987) *J. Virol.* 61:1639-1646; Adam, M.A. and A.D. Miller (1988) *J. Virol.* 62:3802-3806; Dull, T. et al. (1998) *J. Virol.* 72:8463-8471; Zufferey, R. et al. (1998) *J. Virol.* 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4<sup>+</sup> T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) *J. Virol.* 71:7020-7029; Bauer, G. et al. (1997) *Blood* 89:2259-2267; Bonyhadi, M.L. (1997) *J. Virol.* 71:4707-4716; Ranga, U. et al. (1998) *Proc. Natl. Acad. Sci. USA* 95:1201-1206; Su, L. (1997) *Blood* 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding IGSFP to cells which have one or more genetic abnormalities with respect to the expression of IGSFP. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csente, M.E. et al. (1995) *Transplantation* 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) *Annu. Rev. Nutr.* 19:511-544 and Verma, I.M. and N. Somia (1997) *Nature* 389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding IGSFP to target cells which have one or more genetic abnormalities with respect to the expression of IGSFP. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing IGSFP to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) *Exp. Eye Res.* 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) *J. Virol.* 73:519-532 and Xu, H. et al. (1994) *Dev. Biol.* 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding IGSFP to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) *Curr. Opin. Biotechnol.* 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for IGSFP into the alphavirus genome in place of the capsid-coding region results in the production of a large number of IGSFP-coding RNAs and the synthesis of high levels of IGSFP in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) *Virology* 228:74-83). The wide host range of alphaviruses will

allow the introduction of IGSFP into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

15 Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding IGSFP.

20 Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of 25 candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. 30 Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding IGSFP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell

lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterate linkages 5 within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a 10 compound which is effective in altering expression of a polynucleotide encoding IGSFP. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular 15 chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of 20 polynucleotide expression. Thus, in the treatment of disorders associated with increased IGSFP expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding IGSFP may be therapeutically useful, and in the treatment of disorders associated with decreased IGSFP expression or activity, a compound which specifically promotes expression of the polynucleotide encoding IGSFP may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in 25 altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a 30 library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding IGSFP is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an *in vitro* cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding IGSFP are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding IGSFP. The amount of hybridization may be quantified, thus forming the

basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide 5 can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library 10 of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. 15 Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and 20 monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's 25 Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of IGSFP, antibodies to IGSFP, and mimetics, agonists, antagonists, or inhibitors of IGSFP.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, 30 intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-

acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without 5 needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of 10 macromolecules comprising IGSFP or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, IGSFP or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et 15 al. (1999) *Science* 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for 20 administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example IGSFP or fragments thereof, antibodies of IGSFP, and agonists, antagonists or inhibitors of IGSFP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by 25 calculating the ED<sub>50</sub> (the dose therapeutically effective in 50% of the population) or LD<sub>50</sub> (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD<sub>50</sub>/ED<sub>50</sub> ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is 30 preferably within a range of circulating concentrations that includes the ED<sub>50</sub> with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the

subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1  $\mu$ g to 100,000  $\mu$ g, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art.

10 Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

#### DIAGNOSTICS

In another embodiment, antibodies which specifically bind IGSFP may be used for the diagnosis of disorders characterized by expression of IGSFP, or in assays to monitor patients being treated with IGSFP or agonists, antagonists, or inhibitors of IGSFP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for IGSFP include methods which utilize the antibody and a label to detect IGSFP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring IGSFP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of IGSFP expression. Normal or standard values for IGSFP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to IGSFP under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of IGSFP expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

30 In another embodiment of the invention, the polynucleotides encoding IGSFP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of IGSFP may be correlated with

disease. The diagnostic assay may be used to determine absence, presence, and excess expression of IGSFP, and to monitor regulation of IGSFP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding IGSFP or closely related molecules may be used to 5 identify nucleic acid sequences which encode IGSFP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding IGSFP, allelic variants, or related sequences.

10 Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the IGSFP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:7-12 or from genomic sequences including promoters, enhancers, and introns of the IGSFP gene.

15 Means for producing specific hybridization probes for DNAs encoding IGSFP include the cloning of polynucleotide sequences encoding IGSFP or IGSFP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as  $^{32}\text{P}$  or  $^{35}\text{S}$ , or by enzymatic labels, 20 such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

25 Polynucleotide sequences encoding IGSFP may be used for the diagnosis of disorders associated with expression of IGSFP. Examples of such disorders include, but are not limited to, an immune system disorder such as acquired immunodeficiency syndrome (AIDS), X-linked agammaglobulinemia of Bruton, common variable immunodeficiency (CVI), DiGeorge's syndrome (thymic hypoplasia), thymic dysplasia, isolated IgA deficiency, severe combined immunodeficiency disease (SCID), immunodeficiency with thrombocytopenia and eczema (Wiskott-Aldrich syndrome), Chediak-Higashi syndrome, chronic granulomatous diseases, hereditary angioneurotic edema, immunodeficiency associated with Cushing's disease, Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune 30 hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's

syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis,

5      thombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor

10     neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases

15     of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis,

20     inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a developmental disorder such as renal tubular acidosis, anemia,

25     Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and

30     cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a muscle disorder such as cardiomyopathy, myocarditis, Duchenne's muscular dystrophy, Becker's muscular dystrophy, myotonic dystrophy, central core disease, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, infectious myositis,

polymyositis, dermatomyositis, inclusion body myositis, thyrotoxic myopathy, and ethanol myopathy; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including 5 adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, 10 ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. The polynucleotide sequences encoding IGSFP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered IGSFP expression. 15 Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding IGSFP may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide 15 sequences encoding IGSFP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding IGSFP in the sample 20 indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of IGSFP, a normal or standard profile for expression is established. This may be accomplished by 25 combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding IGSFP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values 30 obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the

patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or 5 overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

10 Additional diagnostic uses for oligonucleotides designed from the sequences encoding IGSFP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding IGSFP, or a fragment of a polynucleotide complementary to the polynucleotide encoding IGSFP, and will be employed under optimized conditions for identification of a specific gene or 15 condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding IGSFP may be used to detect single nucleotide polymorphisms (SNPs). SNPs are 20 substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding IGSFP are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the 25 secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSSCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed *in silico* SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual 30 overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the

high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of IGSFP include radiolabeling or biotinylation of nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) *J. Immunol. Methods* 159:235-244; Duplaa, C. et al. (1993) *Anal. Biochem.* 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the 10 polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify 15 genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the 20 activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, IGSFP, fragments of IGSFP, or antibodies specific for IGSFP may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to 25 generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by 30 hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The

resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression *in vivo*, as in the case of a tissue or biopsy sample, or *in vitro*, as in the case of a cell line.

5 Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with *in vitro* model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity

10 (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression

15 provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not

20 necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at <http://www.niehs.nih.gov/oc/news/toxchip.htm>.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

25 In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an

30 untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global

pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating 5 and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an 10 agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for 15 example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

20 A proteomic profile may also be generated using antibodies specific for IGSFP to quantify the levels of IGSFP expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) *Anal. Biochem.* 270:103-111; Mendoza, L.G. et al. (1999) *Biotechniques* 27:778-788). Detection may be performed by a 25 variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and 30 should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) *Electrophoresis* 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid

degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of 5 each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

10 In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the 15 two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 20 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding IGSFP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either 25 coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), 30 yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop

genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding IGSFP on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, IGSFP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between IGSFP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with IGSFP, or fragments thereof, and washed. Bound IGSFP is then detected by methods well known in the art. Purified IGSFP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding IGSFP specifically compete with a test compound for binding IGSFP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with IGSFP.

5 In additional embodiments, the nucleotide sequences which encode IGSFP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding 10 description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications and publications, mentioned above and below, including U.S. Ser. No. 60/249,645, are expressly incorporated by reference herein.

15

## EXAMPLES

### I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA) and shown in Table 4, column 5. Some tissues were 20 homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

25 Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the 30 POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using

the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-5 1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), 10 PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte Genomics, Palo Alto CA), or pINCY (Incyte Genomics), or derivatives thereof. Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 $\alpha$ , DH10B, or 15 ElectroMAX DH10B from Life Technologies.

## II. Isolation of cDNA Clones

15 Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 20 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 25 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSCAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

## III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. 30 Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared

using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the 5 ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, *supra*, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

10 The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and 15 BLOCKS, PRINTS, DOMO, PRODOM; PROTEOME databases with sequences from Homo sapiens, Rattus norvegicus, Mus musculus, Caenorhabditis elegans, Saccharomyces cerevisiae, Schizosaccharomyces pombe, and Candida albicans (Incyte Genomics, Palo Alto CA); and hidden Markov model (HMM)-based protein family databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. 20 (1996) *Curr. Opin. Struct. Biol.* 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using 25 programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently 30 analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, the PROTEOME databases, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San

Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

5 Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where 10 applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

15 The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:7-12. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 4.

#### IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative immunoglobulin superfamily proteins were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). 20 Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum 25 range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode immunoglobulin superfamily proteins, the encoded polypeptides were analyzed by querying against PFAM models for immunoglobulin superfamily proteins. Potential immunoglobulin superfamily proteins were also identified by homology to Incyte cDNA sequences that had been annotated as immunoglobulin superfamily proteins. These selected 30 Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or

public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences 5 using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

## V. Assembly of Genomic Sequence Data with cDNA Sequence Data

### “Stitched” Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene

10 identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a 15 full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals 20 thus identified were then “stitched” together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or 25 genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

### “Stretched” Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST

30 analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate; rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in

Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous 5 genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

#### VI. Chromosomal Mapping of IGSFP Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:7-12 were compared with 10 sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:7-12 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for 15 Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p- 20 arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other 25 resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (<http://www.ncbi.nlm.nih.gov/genemap/>), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

#### VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a 30 gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel (1995) supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related

molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar.

The basis of the search is the product score, which is defined as:

5

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum}\{\text{length(Seq. 1), length(Seq. 2)}\}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding IGSFP are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided

by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding IGSFP. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

### VIII. Extension of IGSFP Encoding Polynucleotides

5 Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30  
10 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

15 High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg<sup>2+</sup>, (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters  
20 for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

25 The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the  
30 concentration of DNA. A 5 µl to 10 µl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates,

digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended 5 clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

10 The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 15 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

20 In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

#### IX. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:7-12 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base 25 pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250  $\mu$ Ci of [ $\gamma$ -<sup>32</sup>P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a 30 SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing  $10^7$  counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate.

5 Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

#### X. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, *supra*), mechanical 10 microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), *supra*).

Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may 15 be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) *Science* 270:467-470; Shalon, D. et al. (1996) *Genome Res.* 6:639-645; Marshall, A. and J. Hodgson (1998) *Nat. Biotechnol.* 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may 20 comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a 25 fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorption and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

30 **Tissue or Cell Sample Preparation**

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)<sup>+</sup> RNA is purified using the oligo-(dT) cellulose method. Each poly(A)<sup>+</sup> RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/ $\mu$ l oligo-(dT) primer (21mer), 1X first

strand buffer, 0.03 units/ $\mu$ l RNase inhibitor, 500  $\mu$ M dATP, 500  $\mu$ M dGTP, 500  $\mu$ M dTTP, 40  $\mu$ M dCTP, 40  $\mu$ M dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)<sup>+</sup> RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)<sup>+</sup> RNAs are synthesized by in vitro transcription 5 from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated 10 using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14  $\mu$ l 5X SSC/0.2% SDS.

#### Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element 15 is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5  $\mu$ g. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope 20 slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

25 Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1  $\mu$ l of the array element DNA, at an average concentration of 100 ng/ $\mu$ l, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). 30 Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

**Hybridization**

Hybridization reactions contain 9  $\mu$ l of sample mixture consisting of 0.2  $\mu$ g each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65° C for 5 minutes and is aliquoted onto the microarray surface and covered with 5 an 1.8 cm<sup>2</sup> coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140  $\mu$ l of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60° C. The arrays are washed for 10 min at 45° C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45° C in a second wash buffer (0.1X SSC), and dried.

**10 Detection**

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide 15 containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, 20 Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

25 The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, 30 are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital

(A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and 5 measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used 10 for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

## XI. Complementary Polynucleotides

Sequences complementary to the IGSFP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring IGSFP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same 15 procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of IGSFP. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the IGSFP-encoding transcript.

## 20 XII. Expression of IGSFP

Expression and purification of IGSFP is achieved using bacterial or virus-based expression systems. For expression of IGSFP in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid 25 promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express IGSFP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of IGSFP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus 30 (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding IGSFP by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to

infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

5 In most expression systems, IGSFP is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from IGSFP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, 10 ch. 10 and 16). Purified IGSFP obtained by these methods can be used directly in the assays shown in Examples XVI and XVII, where applicable.

15

### XIII. Functional Assays

IGSFP function is assessed by expressing the sequences encoding IGSFP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression 20 vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10  $\mu$ g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2  $\mu$ g of an additional plasmid containing sequences encoding a 25 marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate 30 the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light

scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are 5 discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of IGSFP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding IGSFP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using 10 magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding IGSFP and other genes of interest can be analyzed by northern analysis or microarray techniques.

#### XIV. Production of IGSFP Specific Antibodies

15 IGSFP substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

20 Alternatively, the IGSFP amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for 25 selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A 25 peptide synthesizer (Applied Biosystems) using Fmoc chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the 30 oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-IGSFP activity by, for example, binding the peptide or IGSFP to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

#### XV. Purification of Naturally Occurring IGSFP Using Specific Antibodies

Naturally occurring or recombinant IGSFP is substantially purified by immunoaffinity chromatography using antibodies specific for IGSFP. An immunoaffinity column is constructed by

covalently coupling anti-IGSFP antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing IGSFP are passed over the immunoaffinity column, and the column is 5 washed under conditions that allow the preferential absorbance of IGSFP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/IGSFP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and IGSFP is collected.

#### **XVI. Identification of Molecules Which Interact with IGSFP**

10 IGSFP, or biologically active fragments thereof, are labeled with <sup>125</sup>I Bolton-Hunter reagent. (See, e.g., Bolton, A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled IGSFP, washed, and any wells with labeled IGSFP complex are assayed. Data obtained using different concentrations of IGSFP are used to calculate values for the number, affinity, and association of IGSFP with the 15 candidate molecules.

Alternatively, molecules interacting with IGSFP are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) Nature 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

IGSFP may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) 20 which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

#### **XVII. Demonstration of IGSFP Activity**

An assay for IGSFP activity measures the ability of IGSFP to recognize and precipitate 25 antigens from serum. This activity can be measured by the quantitative precipitin reaction (Golub, E.S. et al. (1987) Immunology: A Synthesis, Sinauer Associates, Sunderland MA, pp. 113-115). IGSFP is isotopically labeled using methods known in the art. Various serum concentrations are added to constant amounts of labeled IGSFP. IGSFP-antigen complexes precipitate out of solution and are collected by centrifugation. The amount of precipitable IGSFP-antigen complex is 30 proportional to the amount of radioisotope detected in the precipitate. The amount of precipitable IGSFP-antigen complex is plotted against the serum concentration. For various serum concentrations, a characteristic precipitin curve is obtained, in which the amount of precipitable IGSFP-antigen complex initially increases proportionately with increasing serum concentration, peaks at the

equivalence point, and then decreases proportionately with further increases in serum concentration. Thus, the amount of precipitable IGSFP-antigen complex is a measure of IGSFP activity which is characterized by sensitivity to both limiting and excess quantities of antigen.

Alternatively, an assay for IGSFP activity measures the expression of IGSFP on the cell surface. cDNA encoding IGSFP is transfected into a non-leukocytic cell line. Cell surface proteins are labeled with biotin (de la Fuente, M.A. et.al. (1997) Blood 90:2398-2405). Immunoprecipitations are performed using IGSFP-specific antibodies, and immunoprecipitated samples are analyzed using SDS-PAGE and immunoblotting techniques. The ratio of labeled immunoprecipitant to unlabeled immunoprecipitant is proportional to the amount of IGSFP expressed on the cell surface.

10. Alternatively, an assay for IGSFP activity measures the amount of cell aggregation induced by overexpression of IGSFP. In this assay, cultured cells such as NIH3T3 are transfected with cDNA encoding IGSFP contained within a suitable mammalian expression vector under control of a strong promoter. Cotransfection with cDNA encoding a fluorescent marker protein, such as Green Fluorescent Protein (CLONTECH), is useful for identifying stable transfectants. The amount of cell 15 agglutination, or clumping, associated with transfected cells is compared with that associated with untransfected cells. The amount of cell agglutination is a direct measure of IGSFP activity.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention.

20. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
6471113	1	6471113CD1	7	6471113CB1
7701502	2	7701502CD1	8	7701502CB1
5727379	3	5727379CD1	9	5727379CB1
5831801	4	5831801CD1	10	5831801CB1
7195090	5	7195090CD1	11	7195090CB1
7473826	6	7473826CD1	12	7473826CB1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability Score	GenBank Homolog
1	6471113CD1	g3068592	0.0	Punc [Mus musculus] (Salbaum, J.M. (1998) Mech. Dev. 71:201-204)
2	7701502CD1	g96746556	8.8e-74	sialic acid-binding immunoglobulin-like lectin-8 [Homo sapiens] (Floyd, H. et al. (2000) J. Biol. Chem. 275:861-866)
3	5727379CD1	g9837568	7.1e-43	neuronal leucine-rich repeat protein-3 [Rattus norvegicus]
4	5831801CD1	g312590	3.4e-15	biliary glycoprotein [Mus musculus] (McCuig, K. et al. (1993) Gene 127:173-183)
5	7195090CD1	g97290075	0.0	sdk gene product [Drosophila melanogaster] (Adams, M.D. et al. (2000) Science 287:2185-2195)

Table 3

SEQ ID NO:	Incute Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	6471113CD1	793	S108 S128 S216 S272 S412 S436 S443 S486 S528 S582 S611 S616 S665 S8 T137 T148 T181 T502 T517 T606 T74	N246 N381 N382 N580 N604 N93	Signal peptide: M1-G35 Signal peptide: M1-G35 Transmembrane domain: I623-F641 Immunoglobulin domains: G56-A119, G153-A211, H252-A309, G344-A402, Fibronectin type III domains: P424-S510, A521-T606 Receptor tyrosine kinase BL00790: E439-A464, D487-T517 Score/strength=0.68	SPSCAN HMMER HMMER-PFAM HMMER-PFAM
2	7701502CD1	475	S220 S318 S85 T143 T203 T229 T322 T357 T393 T67 Y321	N100 N355 N364	PUTATIVE NEURONAL CELL ADHESION MOLECULE PUNC PD123980: N604-A783 BASIC FIBROBLAST GROWTH FACTOR RECEPTOR 1 DM01287: A39752 1-814: P244-E421, L16-V118, Q134-V210 P16092 1-821: P244-A420, W17-G124 A60350 1-800: V126-G374, G344-E421 IMMUNOGLOBULIN DM00001 P43146 328-410: P329-Q410 Signal cleavage: M1-A16 Transmembrane domain: A457-M475	BLIMPS-BLOCKS BLAST-PRODOM BLAST-PRODOM BLAST-PRODOM BLAST-DM0

Table 3 (cont.)

SEQ ID NO.	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
2					MYELIN; SCHWANN; STIALOADHESIN; FORM DM03744 [P20138] 1-142; M1-T143 Immunoglobulins and major histocompatibility complex proteins Signature F423-H429	BLAST-DOOM
3	5727379CD1	578	S109 S169 S315 S364 S477 T132 T307 T359 T378 T418 T425 T471 Y234 Y553	N130 N188 N250 N260 N299 N38 N463 N494 N499	Transmembrane domain: M519-W538 Leucine Rich Repeat: S202-F225, A318-Q340, E58-P81, L82-F105, N130-H153, S178-R201, N106-S129, N154-L177, N250-E273, F294-R317 Immunoglobulin domain: G397-A457 Leucine rich repeat N-terminal domain: G27-P56	HMMER
4	5831801CD1	298	S125 S240 S266 S62 T156 T60 T90	N104 N138 N167 N172 N189 N35	Signal peptide: M1-G33 Transmembrane domain: I243-W263 Immunoglobulin domains: G48-Y120, S161-V219 ANTIGEN PRECURSOR SIGNAL IMMUNOGLOBULIN FOLD GLYCOPROTEIN T CELL SURFACE TRANSMEMBRANE: PD010953: I43-V231	HMMER - PFAM
5	7195090CD1	2066	S1036 S1063 S109 S1104 S1173 S1267 S1327 S1368 S1422 S1440 S1501 S1523 S1590 S1620 S1695 S1779 S1788 S192 S1922 S1941 S1952 S1971	N1156 N1207 N145 N1528 N1622 N1641 N1693 N175 N1767 N1914 N1955 N424 N437 N446 N529 N553	Transmembrane domain: W1882-F1901 Fibronectin type III domains: P542-S629, P642-S731, P743-S834, P846-S932, P944-S1034, P1047-S1138, P1150-S1239, P1252-L1338, P1351-S1439, P1452-S1563, P1575-S1664, P1676-Q1765, P1777-S1864	HMMER - PFAM

Table 3 (cont.)

SEQ ID NO:	Incycyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
5			S1989 S1998 S2005 S2029 S243 S398 S478 S652 S66 S853 S861 S865 S956	N656 N695 N756 N889 N898	Immunoglobulin domains: I30-V45, P93-Y134, S182-A238, E275-A334, G368-A429, G462-I523, C824-E903	HMMER-PFAM
			T1055 T1058 T1112 T1247 T126 T1300		Fibronectin type III repeats PRO0014: T860-P869, N672-Y682, V595-Y613, Y1848-E1862	BLIMPS-PRINTS
			T1307 T1312 T1345 T1365			BLAST-DO MO
			T1447 T1507 T1534 T1624		BASIC FIBROBLAST GROWTH FACTOR RECEPTOR 1 DM01287 A39752 1-814: E267-S398, P324-V444, L153-F245	BLAST-DO MO
			T1917 T1923 T2003 T202		IMMUNOGLOBULIN: DM00001 P20241 428-516: P354-V444	BLAST-DO MO
			T222 T227 T25 T345 T357 T364		FIBRONECTIN TYPE III REPEAT: DM00007 S50893 585-685: G624-Y722	BLAST-DO MO
			T386 T495 T632 T697 T713 T816 T827 T841 T900 T994			
6	7473826CD1	333	S310 S69 S89 T101 T164 T230 T316 T325 T35 T5	N125 N160 N256 N261 N33 N65	Signal peptide: M1-A60 IPT/TIG domains (immunoglobulin-like fold shared by Plexins and transcription factors): P131-A211, P216-P298	SPSCAN HMMER-PFAM
					RECEPTOR KINASE TYROSINE PROTEIN PRECURSOR TYROSINEPROTEIN SIGNAL HEPATOCYTE GROWTH ATTBINDING PD003981: S118-N238, T2-P76, P223-V269	BLAST-PRODOM
						e-value=1.3e-10

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
7	6471113CB1	2700	1-627, 1-139, 1711-2116, 410-1502, 1859-2134	7184821H1 (BONRFEC01) 7744009H1 (ADRETUE04) 5048190F8 (PLACFER01) GNN.g8389522_000026_002.edit 1449087H1 (PLACNOT02) GNN.g8389522_000027_002.edit 6471113F8 (PLACFEB01) 5048190R8 (PLACFER01) 71553470V1 2414883H1 (HNT3AZT01)	2152 104 512 410 1614 1 977 796 1966 1395 1953	2700 686 1040 2382 1857 409 1389 1296 2464 1617 2321
8	7701502CB1	2321	1915-1938, 1242-1385, 1-1001, 1988-2063, 1453-1503	71583648V1 71583740V1 71571142V1 7168805H1 (MCLRNC01) 4991695T6 (LIVRTUT11)	1348 1302 1302 537 1	2090 1916 1617 2321 621
9	5727379CB1	2112	2052-2112, 1-456, 1041-1128, 864-1119	1709963F6 (PROSNOT16) FL5727379_g8176790_000004_g315 7804 7065822H1 (BRATNOR01)	674 1 1516	1333 1734 2112
10	5831801CB1	1419	1-262, 1-452, 1123-1364, 1368-1419	6884626H1 (BRAHTDR03) 7091169R8 (BRAUTDR03) 6891983H1 (BRAUTDR03) 7091169F8 (BRAUTDR03)	873 1 800 652	1419 802 1381 1301

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
11	7195090CB1	6374	1007-2395, 2744-6374, 1-144, 490-6119	7674279H2 (NOSETTE01) 7674694J1 (NOSETTE01) 92993996 7262823H1 (UTREPMAC01) 7195090H1 (LUNGFER04) 71185491V1 7634914H1 (SINTDIE01) 56002695J1 7198117R8 (LUNGFER04)	3281 4711 178 2105 1908 5722 2677 3863 243	33386 5268 489 2736 2464 6374 3235 4838 1029
				7634914J1 (SINTDIE01) GNN.g6042167_008.edit 72080251D1 7198133F8 (LUNGFER04) 6879120F8 (LNODNOR03) GBI.g10334962_000008_000011.ed it	3304 1 5056 947 1235 2633 817 1	3966 335 5809 1605 2011 3445 1350 1002
12	7473826CB1	1350	1250-1350, 1-341, 1-982, 6555 760-1063	5001304F8 (PROSTUTR21) FI7473826_g7382420_000006_g600		

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID	Representative Library
7	6471113CB1	HNT2AZS07
8	7701502CB1	EC51N0T02
9	5727379CB1	UTRSN0T12
10	5831801CB1	BRAUTD03
11	7195090CB1	LUNGFER04
12	7473826CB1	PROSTUT21

Table 6

Library	Vector	Library Description
BRAUTD03	PCDNA2.1	This random primed library was constructed using RNA isolated from pooled globus pallidus and substantia innominata tissue removed from a 55-year-old Caucasian female who died from cholangiocarcinoma. Pathology indicated mild meningeal fibrosis predominately over the convexities, scattered axonal spheroids in the white matter of the cingulate cortex and the thalamus, and a few scattered neurofibrillary tangles in the entorhinal cortex and the periaqueductal gray region. Pathology for the associated tumor tissue indicated well-differentiated cholangiocarcinoma of the liver with residual or relapsed tumor. Patient history included cholangiocarcinoma, post-operative Budd-Chiari syndrome, biliary ascites, hydrothorax, dehydration, malnutrition, oliguria and acute renal failure. Previous surgeries included cholecystectomy and resection of 85% of the liver.
EOSINOT02	PSPORT1	Library was constructed using RNA isolated from pooled eosinophils obtained from allergic asthmatic individuals.
HNT2AZS07	PSPORT1	This subtracted library was constructed from RNA isolated from an HNT2 cell line (derived from a human teratocarcinoma that exhibited properties characteristic of a committed neuronal precursor) treated for three days with 0.35 micromolar AZ. The hybridization probe for subtraction was derived from a similarly constructed library from untreated HNT2 cells. 3.08M clones from the AZ-treated library were subjected to three rounds of subtractive hybridization with 3.04M clones from the untreated library. Subtractive hybridization conditions were based on the methodologies of Swaroop et al. (NAR (1991) 19:1954) and Bonaldo et al. (Genome Research (1996) 6:791).
LUNGFER04	PCDNA2.1	This random primed library was constructed using RNA isolated from lung tissue removed from a Caucasian male fetus who died from fetal demise.

Table 6 (cont.)

Library	Vector	Library Description
PROSTUT21	PINCY	Library was constructed using RNA isolated from right prostate tumor tissue removed from a 61-year-old Caucasian male during a radical prostatectomy and regional lymph node excision. Pathology indicated adenocarcinoma, Gleason grade 3+3, forming a predominant mass involving the right side centrally and peripherally. The tumor invaded the right mid-posterior capsule but did not extend beyond it. The surgical margins were positive, including right apex and right bladder base. The patient presented with induration, hyperplasia of the prostate, and elevated prostate specific antigen. Patient history included renal failure, osteoarthritis, left renal artery stenosis, benign hypertension, thrombocytopenia, hyperlipidemia, tobacco and alcohol abuse in remission, and hepatitis C (carrier). Previous surgeries included a total hip replacement. Patient medications included aspirin and Inderal. Family history included benign hypertension in the mother and in the grandparent(s).
UTRSNOT12	PINCY	Library was constructed using RNA isolated from uterine myometrial tissue removed from a 41-year-old Caucasian female during a vaginal hysterectomy with dilation and curettage. The endometrium was secretory and contained fragments of endometrial polyps. Benign endo- and ectocervical mucosa were identified in the endocervix. Pathology for the associated tumor tissue indicated uterine leiomyoma. Patient history included ventral hernia and a benign ovarian neoplasm.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) <i>J. Mol. Biol.</i> 215:403-410; Altschul, S.F. et al. (1997) <i>Nucleic Acids Res.</i> 25:3389-3402.	<i>ESTs</i> : Probability value= 1.0E-8 or less <i>Full Length sequences</i> : Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, tfasta, and search.	Pearson, W.R. and D.J. Lipman (1988) <i>Proc. Natl. Acad Sci. USA</i> 85:2444-2448; Pearson, W.R. (1990) <i>Methods Enzymol.</i> 183:63-98; and Smith, T.F. and M.S. Waterman (1981) <i>Adv. Appl. Math.</i> 2:482-489.	<i>ESTs</i> : fasta E value=1.0E-6 Assembled <i>ESTs</i> : fasta Identity= 95% or greater and Match length=200 bases or greater; fasta E value=1.0E-8 or less <i>Full Length sequences</i> : fasta score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) <i>Nucleic Acids Res.</i> 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) <i>Methods Enzymol.</i> 266:88-105; and Atwood, T.K. et al. (1997) <i>J. Chem. Inf. Comput. Sci.</i> 37:417-424.	Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) <i>J. Mol. Biol.</i> 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) <i>Nucleic Acids Res.</i> 26:320-322; Durbin, R. et al. (1998) <i>Our World View</i> , in a Nutshell, Cambridge Univ. Press, pp. 1-350.	<i>PFAM hits</i> : Probability value= 1.0E-3 or less <i>Signal peptide hits</i> : Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score $\geq$ GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phrls Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Somhammer, E.L. et al. (1998) Proc. Sixth Int'l. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:
  - a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-6,
  - b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-6,
  - c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and
  - d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6.
2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-6.
3. An isolated polynucleotide encoding a polypeptide of claim 1.
4. An isolated polynucleotide encoding a polypeptide of claim 2.
5. An isolated polynucleotide of claim 4 comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:7-12.
6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.
7. A cell transformed with a recombinant polynucleotide of claim 6.
8. A transgenic organism comprising a recombinant polynucleotide of claim 6.
9. A method of producing a polypeptide of claim 1, the method comprising:
  - a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant

polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

b) recovering the polypeptide so expressed.

5 10. A method of claim 9, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-6.

11. An isolated antibody which specifically binds to a polypeptide of claim 1.

10 12. An isolated polynucleotide selected from the group consisting of:

a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:7-12,

b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:7-12,

c) a polynucleotide complementary to a polynucleotide of a),

d) a polynucleotide complementary to a polynucleotide of b), and

e) an RNA equivalent of a)-d).

20 13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.

14. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

25 a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and

30 b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.

16. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

- a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and
- 5 b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

10

18. A composition of claim 17, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-6.

19. A method for treating a disease or condition associated with decreased expression of 15 functional IGSFP, comprising administering to a patient in need of such treatment the composition of claim 17.

20. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

20

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting agonist activity in the sample.

25 21. A composition comprising an agonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.

25

22. A method for treating a disease or condition associated with decreased expression of functional IGSFP, comprising administering to a patient in need of such treatment a composition of claim 21.

30

23. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting antagonist activity in the sample.

24. A composition comprising an antagonist compound identified by a method of claim 23 and a pharmaceutically acceptable excipient.

25. A method for treating a disease or condition associated with overexpression of functional  
5 IGSFP, comprising administering to a patient in need of such treatment a composition of claim 24.

26. A method of screening for a compound that specifically binds to the polypeptide of claim  
1, the method comprising:

- 10 a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and
- b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

27. A method of screening for a compound that modulates the activity of the polypeptide of  
15 claim 1, the method comprising:

- a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,
- b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and
- 20 c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

25  
28. A method of screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

- 30 a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,
- b) detecting altered expression of the target polynucleotide, and
- c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

29. A method of assessing toxicity of a test compound, the method comprising:

- a) treating a biological sample containing nucleic acids with the test compound,
- b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof,
- c) quantifying the amount of hybridization complex, and
- d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

10

15 30. A diagnostic test for a condition or disease associated with the expression of IGSFP in a biological sample, the method comprising:

- a) combining the biological sample with an antibody of claim 11, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and
- b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

20

25 31. The antibody of claim 11, wherein the antibody is:

- a) a chimeric antibody,
- b) a single chain antibody,
- c) a Fab fragment,
- d) a F(ab')<sub>2</sub> fragment, or
- e) a humanized antibody.

30

32. A composition comprising an antibody of claim 11 and an acceptable excipient.

33. A method of diagnosing a condition or disease associated with the expression of IGSFP in a subject, comprising administering to said subject an effective amount of the composition of claim 32.

34. A composition of claim 32, wherein the antibody is labeled.

35. A method of diagnosing a condition or disease associated with the expression of IGSFP in a subject, comprising administering to said subject an effective amount of the composition of claim 34.

5

36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 11, the method comprising:

- a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, or an immunogenic fragment thereof, under conditions to elicit an antibody response,
- b) isolating antibodies from said animal, and
- c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which binds specifically to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-6.

10

37. A polyclonal antibody produced by a method of claim 36.

38. A composition comprising the polyclonal antibody of claim 37 and a suitable carrier.

20

39. A method of making a monoclonal antibody with the specificity of the antibody of claim 11, the method comprising:

- a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, or an immunogenic fragment thereof, under conditions to elicit an antibody response,
- b) isolating antibody producing cells from the animal,
- c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,
- d) culturing the hybridoma cells, and
- e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-6.

25

30

40. A monoclonal antibody produced by a method of claim 39.

41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.
42. The antibody of claim 11, wherein the antibody is produced by screening a Fab expression library.
- 5 43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant immunoglobulin library.
- 10 44. A method of detecting a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-6 in a sample, the method comprising:
  - a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and
  - b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-6 in the sample.
- 15 45. A method of purifying a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-6 from a sample, the method comprising:
  - a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and
  - b) separating the antibody from the sample and obtaining the purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-6.
- 20 46. A microarray wherein at least one element of the microarray is a polynucleotide of claim 13.
- 25 47. A method of generating an expression profile of a sample which contains polynucleotides, the method comprising:
  - a) labeling the polynucleotides of the sample,
  - b) contacting the elements of the microarray of claim 46 with the labeled polynucleotides of the sample under conditions suitable for the formation of a hybridization complex, and
- 30

c) quantifying the expression of the polynucleotides in the sample.

48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.

49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.

10

50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

15

51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to said target polynucleotide.

52. An array of claim 48, which is a microarray.

20

53. An array of claim 48, further comprising said target polynucleotide hybridized to a nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.

54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to said solid substrate.

25

55. An array of claim 48, wherein each distinct physical location on the substrate contains multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical location have the same sequence, and each distinct physical location on the substrate contains nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.

30

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.
59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.
60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.
61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.
62. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:7.
- 10 63. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:8.
64. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:9.
- 15 65. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:10.
66. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:11.
67. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:12.

20

<110> INCYTE GENOMICS, INC.  
BAUGHN, Mariah R.  
LU, Dyung Aina M.  
YUE, Henry  
ELLIOTT, Vicki S.  
THANGAVELU, Kavitha  
RAMKUMAR, Jayalaxmi  
LU, Yan  
LO, Terrence P.  
GURURAJAN, Rajagopal  
GANDHI, Ameena R.  
ARVIZU, Chandra  
YAO, Monique G.

<120> IMMUNOGLOBULIN SUPERFAMILY PROTEINS

<130> PF-0841 PCT

<140> To Be Assigned  
<141> Herewith

<150> 60/249,645  
<151> 2000-11-16

<160> 12  
<170> PERL Program

<210> 1  
<211> 793  
<212> PRT  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 6471113CD1

<400> 1  
Met Ala Val Gln Arg Ala Ala Ser Pro Arg Arg Pro Pro Ala Pro  
1 5 10 15  
Leu Trp Pro Arg Leu Leu Leu Pro Leu Leu Leu Leu Leu Pro  
20 25 30  
Ala Pro Ser Glu Gly Leu Gly His Ser Ala Glu Leu Ala Phe Ala  
35 40 45  
Val Glu Pro Ser Asp Asp Val Ala Val Pro Gly Gln Pro Ile Val  
50 55 60  
Leu Asp Cys Arg Val Glu Gly Thr Pro Pro Val Arg Ile Thr Trp  
65 70 75  
Arg Lys Asn Gly Val Glu Leu Pro Glu Ser Thr His Ser Thr Leu  
80 85 90  
Leu Ala Asn Gly Ser Leu Met Ile Arg His Phe Arg Leu Glu Pro  
95 100 105  
Gly Gly Ser Pro Ser Asp Glu Gly Asp Tyr Glu Cys Val Ala Gln  
110 115 120  
Asn Arg Phe Gly Leu Val Val Ser Arg Lys Ala Arg Ile Gln Ala  
125 130 135  
Ala Thr Met Ser Asp Phe His Val His Pro Gln Ala Thr Val Gly

140	145	150
Glu Glu Gly Gly Val Ala Arg Phe Gln Cys Gln Ile His Gly Leu		
155	160	165
Pro Lys Pro Leu Ile Thr Trp Glu Lys Asn Arg Val Pro Ile Asp		
170	175	180
Thr Asp Asn Glu Arg Tyr Thr Leu Leu Pro Lys Gly Val Leu Gln		
185	190	195
Ile Thr Gly Leu Arg Ala Glu Asp Gly Gly Ile Phe His Cys Val		
200	205	210
Ala Ser Asn Ile Ala Ser Ile Arg Ile Ser His Gly Ala Arg Leu		
215	220	225
Thr Val Ser Gly Ser Gly Ser Gly Ala Tyr Lys Glu Pro Ala Ile		
230	235	240
Leu Val Gly Pro Glu Asn Leu Thr Leu Thr Val His Gln Thr Ala		
245	250	255
Val Leu Glu Cys Val Ala Thr Gly Asn Pro Arg Pro Ile Val Ser		
260	265	270
Trp Ser Arg Leu Asp Gly Arg Pro Ile Gly Val Glu Gly Ile Gln		
275	280	285
Val Leu Gly Thr Gly Asn Leu Ile Ile Ser Asp Val Thr Val Gln		
290	295	300
His Ser Gly Val Tyr Val Cys Ala Ala Asn Arg Pro Gly Thr Arg		
305	310	315
Val Arg Arg Thr Ala Gln Gly Arg Leu Val Val Gln Ala Pro Ala		
320	325	330
Glu Phe Val Gln His Pro Gln Ser Ile Ser Arg Pro Ala Gly Thr		
335	340	345
Thr Ala Met Phe Thr Cys Gln Ala Gln Gly Glu Pro Pro Pro His		
350	355	360
Val Thr Trp Leu Lys Asn Gly Gln Val Leu Gly Pro Gly Gly His		
365	370	375
Val Arg Leu Lys Asn Asn Ser Thr Leu Thr Ile Ser Gly Ile		
380	385	390
Gly Pro Glu Asp Glu Ala Ile Tyr Gln Cys Val Ala Glu Asn Ser		
395	400	405
Ala Gly Ser Ser Gln Ala Ser Ala Arg Leu Thr Val Leu Trp Ala		
410	415	420
Glu Gly Leu Pro Gly Pro Pro Arg Asn Val Arg Ala Val Ser Val		
425	430	435
Ser Ser Thr Glu Val Arg Val Ser Trp Ser Glu Pro Leu Ala Asn		
440	445	450
Thr Lys Glu Ile Ile Gly Tyr Val Leu His Ile Arg Lys Ala Ala		
455	460	465
Asp Pro Pro Glu Leu Glu Tyr Gln Glu Ala Val Ser Lys Ser Thr		
470	475	480
Phe Gln His Leu Val Ser Asp Leu Glu Pro Ser Thr Ala Tyr Ser		
485	490	495
Phe Tyr Ile Lys Ala Tyr Thr Pro Arg Gly Ala Ser Ser Ala Ser		
500	505	510
Val Pro Thr Leu Ala Ser Thr Leu Gly Glu Ala Pro Ala Pro Pro		
515	520	525
Pro Leu Ser Val Arg Val Leu Gly Ser Ser Ser Leu Gln Leu Leu		
530	535	540
Trp Glu Pro Trp Pro Arg Leu Ala Gln His Glu Gly Gly Phe Lys		
545	550	555
Leu Phe Tyr Arg Pro Ala Ser Lys Thr Ser Phe Thr Gly Pro Ile		

560	565	570
Leu Leu Pro Gly Thr Val Ser Ser Tyr Asn Leu Ser Gln Leu Asp		
575	580	585
Pro Thr Ala Val Tyr Glu Val Lys Leu Leu Ala Tyr Asn Gln His		
590	595	600
Gly Asp Gly Asn Ala Thr Val Arg Phe Val Ser Leu Arg Gly Ala		
605	610	615
Ser Glu Arg Thr Gly Ile Val Ile Gly Ile His Ile Gly Val Thr		
620	625	630
Cys Ile Ile Phe Cys Val Leu Phe Leu Leu Phe Gly Gln Arg Gly		
635	640	645
Arg Val Leu Leu Cys Lys Asp Val Glu Asn Gln Leu Ser Pro Pro		
650	655	660
Gln Gly Pro Arg Ser Gln Arg Asp Pro Gly Ile Leu Ala Leu Asn		
665	670	675
Gly Ala Arg Arg Gly Gln Arg Gly Gln Leu Gly Arg Asp Glu Lys		
680	685	690
Arg Val Asp Met Lys Glu Leu Glu Gln Leu Phe Pro Pro Ala Ser		
695	700	705
Ala Ala Gly Gln Pro Asp Pro Arg Pro Thr Asp Pro Ala Ala Pro		
710	715	720
Ala Pro Cys Glu Glu Thr Gln Leu Ser Leu Leu Pro Leu Gln Gly		
725	730	735
Cys Gly Leu Met Glu Gly Lys Thr Thr Glu Ala Lys Thr Thr Glu		
740	745	750
Ala Thr Ala Pro Cys Ala Gly Leu Ala Ala Ala Pro Pro Pro Pro		
755	760	765
Asp Gly Gly Pro Gly Leu Leu Ser Glu Gly Gln Ala Ser Arg Pro		
770	775	780
Ala Ala Ala Arg Val Thr Gln Pro Ala His Ser Glu Gln		
785	790	

&lt;210&gt; 2

&lt;211&gt; 475

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7701502CD1

&lt;400&gt; 2

Met Leu Leu Pro Leu Leu Leu Ser Ser Leu Leu Gly Gly Ser Gln		
1	5	10
Ala Met Asp Gly Arg Phe Trp Ile Arg Val Gln Glu Ser Val Met		
20	25	30
Val Pro Glu Gly Leu Cys Ile Ser Val Pro Cys Ser Phe Ser Tyr		
35	40	45
Pro Arg Gln Asp Trp Thr Gly Ser Thr Pro Ala Tyr Gly Tyr Trp		
50	55	60
Phe Lys Ala Val Thr Glu Thr Thr Lys Gly Ala Pro Val Ala Thr		
65	70	75
Asn His Gln Ser Arg Glu Val Glu Met Ser Thr Arg Gly Arg Phe		
80	85	90
Gln Leu Thr Gly Asp Pro Ala Lys Gly Asn Cys Ser Leu Val Ile		
95	100	105

HLL

Arg Asp Ala Gln Met Gln Asp Glu Ser Gln Tyr Phe Phe Arg Val		
110	115	120
Glu Arg Gly Ser Tyr Val Arg Tyr Asn Phe Met Asn Asp Gly Phe		
125	130	135
Phe Leu Lys Val Thr Ala Leu Thr Gln Lys Pro Asp Val Tyr Ile		
140	145	150
Pro Glu Thr Leu Glu Pro Gly Gln Pro Val Thr Val Ile Cys Val		
155	160	165
Phe Asn Trp Ala Phe Glu Glu Cys Pro Pro Pro Ser Phe Ser Trp		
170	175	180
Thr Gly Ala Ala Leu Ser Ser Gln Gly Thr Lys Pro Thr Thr Ser		
185	190	195
His Phe Ser Val Leu Ser Phe Thr Pro Arg Pro Gln Asp His Asn		
200	205	210
Thr Asp Leu Thr Cys His Val Asp Phe Ser Arg Lys Gly Val Ser		
215	220	225
Val Gln Arg Thr Val Arg Leu Arg Val Ala Tyr Ala Pro Arg Asp		
230	235	240
Leu Val Ile Ser Ile Ser Arg Asp Asn Thr Pro Ala Leu Glu Pro		
245	250	255
Gln Pro Gln Gly Asn Val Pro Tyr Leu Glu Ala Gln Lys Gly Gln		
260	265	270
Phe Leu Arg Leu Leu Cys Ala Ala Asp Ser Gln Pro Pro Ala Thr		
275	280	285
Leu Ser Trp Val Leu Gln Asn Arg Val Leu Ser Ser Ser His Pro		
290	295	300
Trp Gly Pro Arg Pro Leu Gly Leu Glu Leu Pro Gly Val Lys Ala		
305	310	315
Gly Asp Ser Gly Arg Tyr Thr Cys Arg Ala Glu Asn Arg Leu Gly		
320	325	330
Ser Gln Gln Arg Ala Leu Asp Leu Ser Val Gln Tyr Pro Pro Glu		
335	340	345
Asn Leu Arg Val Met Val Ser Gln Ala Asn Arg Thr Val Leu Glu		
350	355	360
Asn Leu Gly Asn Gly Thr Ser Leu Pro Val Leu Glu Gly Gln Ser		
365	370	375
Leu Cys Leu Val Cys Val Thr His Ser Ser Pro Pro Ala Arg Leu		
380	385	390
Ser Trp Thr Gln Arg Gly Gln Val Leu Ser Pro Ser Gln Pro Ser		
395	400	405
Asp Pro Gly Val Leu Glu Leu Pro Arg Val Gln Val Glu His Glu		
410	415	420
Gly Glu Phe Thr Cys His Ala Arg His Pro Leu Gly Ser Gln His		
425	430	435
Val Ser Leu Ser Leu Ser Val His Tyr Lys Lys Gly Leu Ile Ser		
440	445	450
Thr Ala Phe Ser Asn Gly Ala Phe Leu Gly Ile Gly Ile Thr Ala		
455	460	465
Leu Leu Phe Leu Cys Leu Ala Leu Ile Met		
470	475	

&lt;210&gt; 3

&lt;211&gt; 578

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 5727379CD1

&lt;400&gt; 3

Met Leu His Thr Ala Ile Ser Cys Trp Gln Pro Phe Leu Gly Leu  
1 5 10 15  
Ala Val Val Leu Ile Phe Met Gly Ser Thr Ile Gly Cys Pro Ala  
20 25 30  
Arg Cys Glu Cys Ser Ala Gln Asn Lys Ser Val Ser Cys His Arg  
35 40 45  
Arg Arg Leu Ile Ala Ile Pro Glu Gly Ile Pro Ile Glu Thr Lys  
50 55 60  
Ile Leu Asp Leu Ser Lys Asn Arg Leu Lys Ser Val Asn Pro Glu  
65 70 75  
Glu Phe Ile Ser Tyr Pro Leu Leu Glu Ile Asp Leu Ser Asp  
80 85 90  
Asn Ile Ile Ala Asn Val Glu Pro Gly Ala Phe Asn Asn Leu Phe  
95 100 105  
Asn Leu Arg Ser Leu Arg Leu Lys Gly Asn Arg Leu Lys Leu Val  
110 115 120  
Pro Leu Gly Val Phe Thr Gly Leu Ser Asn Leu Thr Lys Leu Asp  
125 130 135  
Ile Ser Glu Asn Lys Ile Val Ile Leu Leu Asp Tyr Met Phe Gln  
140 145 150  
Asp Leu His Asn Leu Lys Ser Leu Glu Val Gly Asp Asn Asp Leu  
155 160 165  
Val Tyr Ile Ser His Arg Ala Phe Ser Gly Leu Leu Ser Leu Glu  
170 175 180  
Gln Leu Thr Leu Glu Lys Cys Asn Leu Thr Ala Val Pro Thr Glu  
185 190 195  
Ala Leu Ser His Leu Arg Ser Leu Ile Ser Leu His Leu Lys His  
200 205 210  
Leu Asn Ile Asn Asn Met Pro Val Tyr Ala Phe Lys Arg Leu Phe  
215 220 225  
His Leu Lys His Leu Glu Ile Asp Tyr Trp Pro Leu Leu Asp Met  
230 235 240  
Met Pro Ala Asn Ser Leu Tyr Gly Leu Asn Leu Thr Ser Leu Ser  
245 250 255  
Val Thr Asn Thr Asn Leu Ser Thr Val Pro Phe Leu Ala Phe Lys  
260 265 270  
His Leu Glu Leu His Ile Val Gly Ala Gln Leu Arg Thr Ile Glu  
275 280 285  
Pro His Ser Phe Gln Gly Leu Arg Phe Leu Arg Val Leu Asn Val  
290 295 300  
Ser Gln Asn Leu Leu Glu Thr Leu Glu Asn Val Phe Ser Ser  
305 310 315  
Pro Arg Ala Leu Glu Val Leu Ser Ile Asn Asn Asn Pro Leu Ala  
320 325 330  
Cys Asp Cys Arg Leu Leu Trp Ile Leu Gln Arg Gln Pro Thr Leu  
335 340 345  
Gln Phe Gly Gly Gln Gln Pro Met Cys Ala Gly Pro Asp Thr Ile  
350 355 360  
Arg Glu Arg Ser Phe Lys Asp Phe His Ser Thr Ala Leu Ser Phe  
365 370 375  
Tyr Phe Thr Cys Lys Pro Lys Ile Arg Glu Lys Lys Leu Gln

380	385	390
His Leu Leu Val Asp Glu Gly Gln Thr Val Gln Leu Glu Cys Ser		
395	400	405
Ala Asp Gly Asp Pro Gln Pro Val Ile Ser Trp Val Thr Pro Arg		
410	415	420
Arg Arg Phe Ile Thr Thr Lys Ser Asn Gly Arg Ala Thr Val Leu		
425	430	435
Gly Asp Gly Thr Leu Glu Ile Arg Phe Ala Gln Asp Gln Asp Ser		
440	445	450
Gly Met Tyr Val Cys Ile Ala Ser Asn Ala Ala Gly Asn Asp Thr		
455	460	465
Phe Thr Ala Ser Leu Thr Val Lys Gly Phe Ala Ser Asp Arg Phe		
470	475	480
Leu Tyr Ala Asn Arg Thr Pro Met Tyr Met Thr Asp Ser Asn Asp		
485	490	495
Thr Ile Ser Asn Gly Thr Asn Ala Asn Thr Phe Ser Leu Asp Leu		
500	505	510
Lys Thr Ile Leu Val Ser Thr Ala Met Gly Cys Phe Thr Phe Leu		
515	520	525
Gly Val Val Leu Phe Cys Phe Leu Leu Leu Phe Val Trp Ser Arg		
530	535	540
Gly Lys Gly Lys His Lys Asn Ser Ile Asp Leu Glu Tyr Val Pro		
545	550	555
Arg Lys Asn Asn Gly Ala Val Val Glu Gly Glu Val Ala Gly Pro		
560	565	570
Arg Arg Phe Asn Met Lys Met Ile		
575		

&lt;210&gt; 4

&lt;211&gt; 298

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 5831801CD1

&lt;400&gt; 4

Met Lys Arg Glu Arg Gly Ala Leu Ser Arg Ala Ser Arg Ala Leu		
1	5	10
Arg Leu Ala Pro Phe Val Tyr Leu Leu Ile Gln Thr Asp Pro		
20	25	30
Leu Glu Gly Val Asn Ile Thr Ser Pro Val Arg Leu Ile His Gly		
35	40	45
Thr Val Gly Lys Ser Ala Leu Leu Ser Val Gln Tyr Ser Ser Thr		
50	55	60
Ser Ser Asp Arg Pro Val Val Lys Trp Gln Leu Lys Arg Asp Lys		
65	70	75
Pro Val Thr Val Val Gln Ser Ile Gly Thr Glu Val Ile Gly Thr		
80	85	90
Leu Arg Pro Asp Tyr Arg Asp Arg Ile Arg Leu Phe Glu Asn Gly		
95	100	105
Ser Leu Leu Leu Ser Asp Leu Gln Leu Ala Asp Glu Gly Thr Tyr		
110	115	120
Glu Val Glu Ile Ser Ile Thr Asp Asp Thr Phe Thr Gly Glu Lys		
125	130	135

Thr Ile Asn Leu Thr Val Asp Val Pro Ile Ser Arg Pro Gln Val  
 140 145 150  
 Leu Val Ala Ser Thr Thr Val Leu Glu Leu Ser Glu Ala Phe Thr  
 155 160 165  
 Leu Asn Cys Ser His Glu Asn Gly Thr Lys Pro Ser Tyr Thr Trp  
 170 175 180  
 Leu Lys Asp Gly Lys Pro Leu Leu Asn Asp Ser Arg Met Leu Leu  
 185 190 195  
 Ser Pro Asp Gln Lys Val Leu Thr Ile Thr Arg Val Leu Met Glu  
 200 205 210  
 Asp Asp Asp Leu Tyr Ser Cys Met Val Glu Asn Pro Ile Ser Gln  
 215 220 225  
 Gly Arg Ser Leu Pro Val Lys Ile Thr Val Tyr Arg Arg Ser Ser  
 230 235 240  
 Leu Tyr Ile Ile Leu Ser Thr Gly Gly Ile Phe Leu Leu Val Thr  
 245 250 255  
 Leu Val Thr Val Cys Ala Cys Trp Lys Pro Ser Lys Arg Lys Gln  
 260 265 270  
 Lys Lys Leu Glu Lys Gln Asn Ser Leu Glu Tyr Met Asp Gln Asn  
 275 280 285  
 Asp Asp Arg Leu Lys Pro Ala Pro Lys Asp His Ser Pro  
 290 295

&lt;210&gt; 5

&lt;211&gt; 2066

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7195090CD1

&lt;400&gt; 5

Met Asp Glu Ala Gly Asn His His Ser Gln Gln Thr Asn Thr Arg  
 1 5 10 15  
 Thr Lys Asn Gln Thr Pro His Val Leu Thr His Arg Tyr Ile Ile  
 20 25 30  
 Pro Ser Leu Gln Lys Leu Asp Ala Gly Phe Tyr Arg Cys Val Val  
 35 40 45  
 Arg Asn Arg Met Gly Ala Leu Leu Gln Arg Lys Ser Glu Val Gln  
 50 55 60  
 Val Ala Tyr Met Gly Ser Phe Met Asp Thr Asp Gln Arg Lys Thr  
 65 70 75  
 Val Ser Gln Gly Arg Ala Ala Ile Leu Asn Leu Leu Pro Ile Thr  
 80 85 90  
 Ser Tyr Pro Arg Pro Gln Val Thr Trp Phe Arg Glu Gly His Lys  
 95 100 105  
 Ile Ile Pro Ser Asn Arg Ile Ala Ile Thr Leu Glu Asn Gln Leu  
 110 115 120  
 Val Ile Leu Ala Thr Thr Ser Asp Ala Gly Ala Tyr Tyr Val  
 125 130 135  
 Gln Ala Val Asn Glu Lys Asn Gly Glu Asn Lys Thr Ser Pro Phe  
 140 145 150  
 Ile His Leu Ser Ile Ala Arg Asp Val Gly Thr Pro Glu Thr Met  
 155 160 165  
 Ala Pro Thr Ile Val Val Pro Pro Gly Asn Arg Ser Val Val Ala

170	175	180
Gly Ser Ser Glu Asn Thr Leu Glu Cys	Ile Ala Ser Ala Arg Pro	
185	190	195
Val Glu Asp Leu Ser Val Thr Trp Lys	Arg Asn Gly Val Arg Ile	
200	205	210
Thr Ser Gly Leu His Ser Phe Gly Arg	Arg Leu Thr Ile Ser Asn	
215	220	225
Pro Thr Ser Ala Asp Thr Gly Pro Tyr	Val Cys Glu Ala Ala Leu	
230	235	240
Pro Gly Ser Ala Phe Glu Pro Ala Arg	Ala Thr Ala Phe Leu Phe	
245	250	255
Ile Ile Glu Pro Pro Tyr Phe Thr Ala	Glu Pro Glu Ser Arg Ile	
260	265	270
Ser Ala Glu Val Glu Glu Thr Val Asp	Ile Gly Cys Gln Ala Met	
275	280	285
Gly Val Pro Leu Pro Thr Leu Gln Trp	Tyr Lys Asp Ala Ile Ser	
290	295	300
Ile Ser Arg Leu Gln Asn Pro Arg Tyr	Lys Val Leu Ala Ser Gly	
305	310	315
Gly Leu Arg Ile Gln Lys Leu Arg Pro	Glu Asp Ser Gly Ile Phe	
320	325	330
Gln Cys Phe Ala Ser Asn Glu Gly	Gly Glu Ile Gln Thr His Thr	
335	340	345
Tyr Leu Asp Val Thr Asn Ile Ala Pro	Val Phe Thr Gln Arg Pro	
350	355	360
Val Asp Thr Thr Val Thr Asp Gly Met	Thr Ala Ile Leu Arg Cys	
365	370	375
Glu Val Ser Gly Ala Pro Lys Pro Ala	Ile Thr Trp Lys Arg Glu	
380	385	390
Asn His Ile Leu Ala Ser Gly Ser Val	Arg Ile Pro Arg Phe Met	
395	400	405
Leu Leu Glu Ser Gly Leu Gln Ile Ala	Pro Val Phe Ile Gln	
410	415	420
Asp Ala Gly Asn Tyr Thr Cys Tyr Ala	Ala Asn Thr Glu Gly Ser	
425	430	435
Leu Asn Ala Ser Ala Thr Leu Thr Val	Trp Asn Arg Thr Ser Ile	
440	445	450
Val His Pro Pro Glu Asp His Val Val	Ile Lys Gly Thr Thr Ala	
455	460	465
Thr Leu His Cys Gly Ala Thr His Asp	Pro Arg Val Ser Leu Arg	
470	475	480
Tyr Val Trp Lys Lys Asp Asn Val Ala	Leu Thr Pro Ser Ser Thr	
485	490	495
Ser Arg Ile Val Val Glu Lys Asp Gly	Ser Leu Leu Ile Ser Gln	
500	505	510
Thr Trp Ser Gly Asp Ile Gly Asp Tyr	Ser Cys Glu Ile Val Ser	
515	520	525
Glu Gly Gly Asn Asp Ser Arg Met Ala	Arg Leu Glu Val Ile Glu	
530	535	540
Leu Pro His Ser Pro Gln Asn Leu Leu	Val Ser Pro Asn Ser Ser	
545	550	555
His Ser His Ala Val Val Leu Ser Trp	Val Arg Pro Phe Asp Gly	
560	565	570
Asn Ser Pro Ile Leu Tyr Tyr Ile Val	Glu Leu Ser Glu Asn Asn	
575	580	585
Ser Pro Trp Lys Val His Leu Ser Asn	Val Gly Pro Glu Met Thr	

590	595	600
Gly Val Thr Val Ser Gly Leu Thr Pro Ala Arg Thr Tyr Gln Phe		
605	610	615
Arg Val Cys Ala Val Asn Glu Val Gly Arg Gly Gln Tyr Ser Ala		
620	625	630
Glu Thr Ser Arg Leu Met Leu Pro Glu Glu Pro Pro Ser Ala Pro		
635	640	645
Pro Lys Asn Ile Val Ala Ser Gly Arg Thr Asn Gln Ser Ile Met		
650	655	660
Val Gln Trp Gln Pro Pro Pro Glu Thr Glu His Asn Gly Val Leu		
665	670	675
Arg Gly Tyr Ile Leu Arg Tyr Arg Leu Ala Gly Leu Pro Gly Glu		
680	685	690
Tyr Gln Gln Arg Asn Ile Thr Ser Pro Glu Val Asn Tyr Cys Leu		
695	700	705
Val Thr Asp Leu Ile Ile Trp Thr Gln Tyr Glu Ile Gln Val Ala		
710	715	720
Ala Tyr Asn Gly Ala Gly Leu Gly Val Phe Ser Arg Ala Val Thr		
725	730	735
Glu Tyr Thr Leu Gln Gly Val Pro Thr Ala Pro Pro Gln Asn Val		
740	745	750
Gln Thr Glu Ala Val Asn Ser Thr Thr Ile Gln Phe Leu Trp Asn		
755	760	765
Pro Pro Pro Gln Gln Phe Ile Asn Gly Ile Asn Gln Gly Tyr Lys		
770	775	780
Leu Leu Ala Trp Pro Ala Asp Ala Pro Glu Ala Val Thr Val Val		
785	790	795
Thr Ile Ala Pro Asp Phe His Gly Val His His Gly His Ile Thr		
800	805	810
Asn Leu Lys Lys Phe Thr Ala Tyr Phe Thr Ser Val Leu Cys Phe		
815	820	825
Thr Thr Pro Gly Asp Gly Pro Pro Ser Thr Pro Gln Leu Val Trp		
830	835	840
Thr Gln Glu Asp Lys Pro Gly Ala Val Gly His Leu Ser Phe Thr		
845	850	855
Glu Ile Leu Asp Thr Ser Leu Lys Val Ser Trp Gln Glu Pro Leu		
860	865	870
Glu Lys Asn Gly Ile Ile Thr Gly Tyr Gln Ile Ser Trp Glu Val		
875	880	885
Tyr Gly Arg Asn Asn Ser Arg Leu Thr His Thr Leu Asn Ser Thr		
890	895	900
Thr His Glu Tyr Lys Ile Gln Gly Leu Ser Ser Leu Thr Thr Tyr		
905	910	915
Thr Ile Asp Val Ala Ala Val Thr Ala Val Gly Thr Gly Leu Val		
920	925	930
Thr Ser Ser Thr Ile Ser Ser Gly Val Pro Pro Asp Leu Pro Gly		
935	940	945
Ala Pro Ser Asn Leu Val Ile Ser Asn Ile Ser Pro Arg Ser Ala		
950	955	960
Thr Leu Gln Phe Arg Pro Gly Tyr Asp Gly Lys Thr Ser Ile Ser		
965	970	975
Arg Trp Ile Val Glu Gly Gln Val Gly Ala Ile Gly Asp Glu Glu		
980	985	990
Glu Trp Val Thr Leu Tyr Glu Glu Glu Asn Glu Pro Asp Ala Gln		
995	1000	1005
Met Leu Glu Ile Pro Asn Leu Thr Pro Tyr Thr His Tyr Arg Phe		

1010	1015	1020
Arg Met Lys Gln Val Asn Ile Val Gly Pro Ser Pro Tyr Ser Pro		
1025	1030	1035
Ser Ser Arg Val Ile Gln Thr Leu Gln Ala Pro Pro Asp Val Ala		
1040	1045	1050
Pro Thr Ser Val Thr Val Arg Thr Ala Ser Glu Thr Ser Leu Arg		
1055	1060	1065
Leu Arg Trp Val Pro Leu Pro Asp Ser Gln Tyr Asn Gly Asn Pro		
1070	1075	1080
Glu Ser Val Gly Tyr Arg Ile Lys Tyr Trp Arg Ser Asp Leu Gln		
1085	1090	1095
Ser Ser Ala Val Ala Gln Val Val Ser Asp Arg Leu Glu Arg Glu		
1100	1105	1110
Phe Thr Ile Glu Leu Glu Glu Trp Met Glu Tyr Glu Leu Gln		
1115	1120	1125
Met Gln Ala Phe Asn Ala Val Gly Ala Gly Pro Trp Ser Glu Val		
1130	1135	1140
Val Arg Gly Arg Thr Arg Glu Ser Val Pro Ser Ala Ala Pro Glu		
1145	1150	1155
Asn Val Ser Ala Glu Ala Val Ser Ser Thr Gln Ile Leu Leu Thr		
1160	1165	1170
Trp Thr Ser Val Pro Glu Gln Asp Gln Asn Gly Leu Ile Leu Gly		
1175	1180	1185
Tyr Lys Ile Leu Phe Arg Ala Lys Asp Leu Asp Pro Glu Pro Arg		
1190	1195	1200
Ser His Ile Val Arg Gly Asn His Thr Gln Ser Ala Leu Leu Ala		
1205	1210	1215
Gly Leu Arg Lys Phe Val Leu Tyr Glu Leu Gln Val Leu Ala Phe		
1220	1225	1230
Thr Arg Ile Gly Asn Gly Val Pro Ser Thr Pro Leu Ile Leu Glu		
1235	1240	1245
Arg Thr Lys Asp Asp Ala Pro Gly Pro Pro Val Arg Leu Val Phe		
1250	1255	1260
Pro Glu Val Arg Leu Thr Ser Val Arg Ile Val Trp Gln Pro Pro		
1265	1270	1275
Glu Glu Pro Asn Gly Ile Ile Leu Gly Tyr Gln Ile Ala Tyr Arg		
1280	1285	1290
Leu Ala Ser Ser Pro His Thr Phe Thr Thr Val Glu Val Gly		
1295	1300	1305
Ala Thr Val Arg Gln Phe Thr Ala Thr Asp Leu Ala Pro Glu Ser		
1310	1315	1320
Ala Tyr Ile Phe Arg Leu Ser Ala Lys Thr Arg Gln Gly Trp Gly		
1325	1330	1335
Glu Pro Leu Glu Ala Thr Val Ile Thr Thr Glu Lys Arg Glu Arg		
1340	1345	1350
Pro Ala Pro Pro Arg Glu Leu Leu Val Pro Gln Ala Glu Val Thr		
1355	1360	1365
Ala Arg Ser Leu Arg Leu Gln Trp Val Pro Gly Ser Asp Gly Ala		
1370	1375	1380
Ser Pro Ile Arg Tyr Phe Thr Met Gln Val Arg Glu Leu Pro Arg		
1385	1390	1395
Gly Glu Trp Gln Thr Tyr Ser Ser Ile Ser His Glu Ala Thr		
1400	1405	1410
Ala Cys Val Val Asp Arg Leu Arg Pro Phe Thr Ser Tyr Lys Leu		
1415	1420	1425
Arg Leu Lys Ala Thr Asn Asp Ile Gly Asp Ser Asp Phe Ser Ser		

1430	1435	1440
Glu Thr Glu Ala Val Thr Thr Leu Gln Asp Val Pro Gly Glu Pro		
1445	1450	1455
Pro Gly Ser Val Ser Ala Thr Pro His Thr Thr Ser Ser Val Leu		
1460	1465	1470
Ile Gln Trp Gln Pro Pro Arg Asp Glu Ser Leu Asn Gly Leu Leu		
1475	1480	1485
Gln Gly Tyr Arg Ile Tyr Tyr Arg Glu Leu Glu Tyr Glu Ala Gly		
1490	1495	1500
Ser Gly Thr Glu Ala Lys Thr Leu Lys Asn Pro Ile Ala Leu His		
1505	1510	1515
Ala Glu Leu Thr Ala Gln Ser Ser Phe Lys Thr Val Asn Ser Ser		
1520	1525	1530
Ser Thr Ser Thr Met Cys Glu Leu Thr His Leu Lys Lys Tyr Arg		
1535	1540	1545
Arg Tyr Glu Val Ile Met Thr Ala Tyr Asn Ile Ile Gly Glu Ser		
1550	1555	1560
Pro Ala Ser Ala Pro Val Glu Val Phe Val Gly Glu Ala Ala Pro		
1565	1570	1575
Ala Met Ala Pro Gln Asn Val Gln Val Thr Pro Leu Thr Ala Ser		
1580	1585	1590
Gln Leu Glu Val Met Trp Asp Pro Pro Pro Glu Ser Gln Asn		
1595	1600	1605
Gly Asn Ile Gln Gly Tyr Lys Ile Tyr Tyr Trp Glu Ala Asp Ser		
1610	1615	1620
Gln Asn Glu Thr Glu Lys Met Lys Val Leu Phe Leu Pro Glu Pro		
1625	1630	1635
Val Val Arg Leu Lys Asn Leu Thr Ser His Thr Lys Tyr Leu Val		
1640	1645	1650
Ser Ile Ser Ala Phe Asn Ala Ala Gly Asp Gly Pro Lys Ser Asp		
1655	1660	1665
Pro Gln Gln Gly Arg Thr His Gln Ala Ala Pro Gly Ala Pro Ser		
1670	1675	1680
Phe Leu Ala Phe Ser Glu Ile Thr Ser Thr Thr Leu Asn Val Ser		
1685	1690	1695
Trp Gly Glu Pro Ala Ala Asn Gly Ile Leu Gln Gly Tyr Arg		
1700	1705	1710
Val Val Tyr Glu Pro Leu Ala Pro Val Gln Gly Val Ser Lys Val		
1715	1720	1725
Val Thr Val Glu Val Arg Gly Asn Trp Gln Arg Trp Leu Lys Val		
1730	1735	1740
Arg Asp Leu Thr Lys Gly Val Thr Tyr Phe Phe Arg Val Gln Ala		
1745	1750	1755
Arg Thr Ile Thr Tyr Gly Pro Glu Leu Gln Ala Asn Ile Thr Ala		
1760	1765	1770
Gly Pro Ala Glu Gly Ser Pro Gly Ser Pro Arg Asp Val Leu Val		
1775	1780	1785
Thr Lys Ser Ala Ser Glu Leu Thr Leu Gln Trp Thr Glu Gly His		
1790	1795	1800
Ser Gly Asp Thr Pro Thr Thr Gly Tyr Val Ile Glu Ala Arg Pro		
1805	1810	1815
Ser Asp Glu Gly Leu Trp Asp Met Phe Val Lys Asp Ile Pro Arg		
1820	1825	1830
Ser Ala Thr Ser Tyr Thr Leu Ser Leu Asp Lys Leu Arg Gln Gly		
1835	1840	1845
Val Thr Tyr Glu Phe Arg Val Val Ala Val Asn Glu Ala Gly Tyr		

1850	1855	1860
Gly Glu Pro Ser Asn Pro Ser Thr Ala Val Ser Ala Gln Val Glu		
1865	1870	1875
Ala Pro Phe Tyr Glu Glu Trp Trp Phe Leu Leu Val Met Ala Leu		
1880	1885	1890
Ser Ser Leu Ile Val Ile Leu Leu Val Val Phe Ala Leu Val Leu		
1895	1900	1905
His Gly Gln Asn Lys Lys Tyr Lys Asn Cys Ser Thr Gly Lys Gly		
1910	1915	1920
Ile Ser Thr Met Glu Glu Ser Val Thr Leu Asp Asn Gly Gly Phe		
1925	1930	1935
Ala Ala Leu Glu Leu Ser Ser Arg His Leu Asn Val Lys Ser Thr		
1940	1945	1950
Phe Ser Lys Lys Asn Gly Thr Arg Ser Pro Pro Arg Pro Ser Pro		
1955	1960	1965
Gly Gly Leu His Tyr Ser Asp Glu Asp Ile Cys Asn Lys Tyr Asn		
1970	1975	1980
Gly Ala Val Leu Thr Glu Ser Val Ser Leu Lys Glu Lys Ser Ala		
1985	1990	1995
Asp Ala Ser Glu Ser Glu Ala Thr Asp Ser Asp Tyr Glu Asp Ala		
2000	2005	2010
Leu Pro Lys His Ser Phe Val Asn His Tyr Met Ser Asp Pro Thr		
2015	2020	2025
Tyr Tyr Asn Ser Trp Lys Arg Arg Ala Gln Gly Arg Ala Pro Ala		
2030	2035	2040
Pro His Ser Val Ala Ile Leu Leu Thr Ser Asn Pro Ser Ala Tyr		
2045	2050	2055
Leu Ser Val Ala Pro Arg Gly Ser Ala Ser Trp		
2060	2065	

&lt;210&gt; 6

&lt;211&gt; 333

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7473826CD1

&lt;400&gt; 6

Met Thr Ser Val Thr Ile Lys Gly Ser Gly Phe Ala Val Ser Ser		
1	5	10
Ala Gly Val Lys Val Leu Met Gly His Phe Pro Cys Lys Val Leu		
20	25	30
Ser Val Asn Tyr Thr Ala Ile Glu Cys Glu Thr Ser Pro Ala Ala		
35	40	45
Gln Gln Leu Val Asp Val Asp Leu Leu Ile His Gly Val Pro Ala		
50	55	60
Gln Cys Gln Gly Asn Cys Thr Phe Ser Tyr Leu Glu Ser Ile Thr		
65	70	75
Pro Tyr Ile Thr Gly Val Phe Pro Asn Ser Val Ile Gly Ser Val		
80	85	90
Lys Val Leu Ile Glu Gly Glu Leu Gly Thr Val Leu Glu Asp		
95	100	105
Ile Phe Gly His His Ser Val Ser Val Val Val Gly Ser Lys Gly		
110	115	120

Leu Ala Leu Gly Asn Leu Thr Val Ser Ser Pro Pro Val Ala Ser  
                  125                 130                 135  
 Leu Ser Pro Thr Ser Gly Ser Ile Gly Gly Gly Thr Thr Leu Val  
                  140                 145                 150  
 Ile Thr Gly Asn Gly Phe Tyr Pro Gly Asn Thr Thr Val Thr Ile  
                  155                 160                 165  
 Gly Asp Glu Pro Cys Gln Ile Ile Ser Ile Asn Pro Asn Glu Val  
                  170                 175                 180  
 Tyr Cys Arg Thr Pro Ala Gly Thr Thr Gly Met Val Asp Val Lys  
                  185                 190                 195  
 Ile Phe Val Asn Thr Ile Ala Tyr Pro Pro Leu Leu Phe Thr Tyr  
                  200                 205                 210  
 Ala Leu Glu Asp Thr Pro Phe Leu Arg Gly Ile Ile Pro Ser Arg  
                  215                 220                 225  
 Gly Pro Pro Gly Thr Glu Ile Glu Ile Thr Gly Ser Asn Phe Gly  
                  230                 235                 240  
 Phe Glu Ile Leu Glu Ile Ser Val Met Ile Asn Asn Ile Gln Cys  
                  245                 250                 255  
 Asn Val Thr Met Ala Asn Asp Ser Val Val Gln Cys Ile Val Gly  
                  260                 265                 270  
 Asp His Ala Gly Gly Thr Phe Pro Val Met Met His His Lys Thr  
                  275                 280                 285  
 Lys Gly Ser Ala Thr Ser Thr Val Val Phe Glu Tyr Pro Leu Asn  
                  290                 295                 300  
 Ile Gln Asn Ile Asn Pro Ser Gln Gly Ser His Lys Tyr Ala Arg  
                  305                 310                 315  
 Thr Cys Arg Tyr Phe Val Ser Val Asn Thr Tyr Leu Glu Thr Leu  
                  320                 325                 330  
 Phe Arg Lys

&lt;210&gt; 7

&lt;211&gt; 2700

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 6471113CB1

&lt;400&gt; 7

atggctgtgc agcgccgcgc gtctccgcgc cgcccccccc cccccgtctg gccccggctc 60  
 ctgctgccgc tgctgttgcgctgctgctgccc gcgcgcagcg agggtcttgg ccactctgtct 120  
 gaactggcat ttgctgtggat gccaaggatgat gatgttggcg tccccgggca gcctatagtg 180  
 ctggactgca ggggtggaggg gacccctcca gtgcgaatca cctggaggaa gaatggggta 240  
 gagctgcccag agagtaccca ctccaccttg ctggccaaatg ggtcccttgat gatccgtcac 300  
 ttcaggctgg agccgggagg cagcccttcg gatgaagggtg actatgagtg tttggcccaag 360  
 aaccgctttg ggctgggttggat cagccggaaag gtcgcatecc aagctgcaac catgtcggac 420  
 ttccacgtgc atccccccaggc caccgtgggt gaggagggtg gtgtggcccg cttccagtgc 480  
 caaatccatg ggcttcccaa accccctgtat acttggggaga agaacagagt cccaaattgac 540  
 acggacaatg agaggatcac attgtgtccc aagggggtcc tgcagatcac aggacttcga 600  
 gctgaggacg gtggcatctt coactgtgtg gcctcaaaaca tcgcaggatcgatccggatcagc 660  
 cacggggcca ggctcactgt gtccaggctcg ggctctgggg cctacaagga gccagccatc 720  
 ctcgtggggc cttggatgggttggatgggtcg tgatgtgtc 780  
 gccacggggca accccgcgc cattgtgtcc tggagccgc tggatgggtcg ccctatcggt 840  
 gtggaggggca tccagggtgttggatgggtcg gggcacagga aacctcatca tctcagacgt gacgggtccag 900

cactctggcg tctacgtctg tgcagccaaac agacctggca cccgggtgag gagaacggca 960  
 caggccggc tggtgtgca agccccagct gagtttgcc agcatccccca gtccatctcc 1020  
 aggccagctg ggaccacagc catgttccacc tgccaaagccc agggtgagcc accgcctcat 1080  
 gtcacgtggc tgaaaaatgg acaggtctg gggccaggag gccacgttag gctcaagaat 1140  
 aacaacagca cactgaccat ttctgaaatc ggtctgtagg atgaagccat ttatcagtgt 1200  
 gtggccgaga acagtgcggg ctcatcacag gccagtggca ggctgaccgt actgtgggct 1260  
 gaggggctcc cggggctcc cgcgaatgtg cgggcgtct ctgtgtcttc cactgagggt 1320  
 cgtgtgtct ggagtggcc getggccaaac accaaggaga tcatcgctca cgtcctgcac 1380  
 atcaggaagg ctgctgaccc accggagctg gatgtatcagg aggcaagttag caagagcacc 1440  
 tttcagccacc tggtcagcga cctggagccc tccacagct acagtttcta catcaaggcc 1500  
 tacacaccaa gggggccag ctcagccctc gtgcggccccc tagtagcact cctgggtgaa 1560  
 gcccctgccc caccggccact gtcagtgcga gtctggca gctctctt gcagctgctg 1620  
 tgggagccctt ggccccggct ggcccagcac gagggccgct tcaagctgtt ttaccggccca 1680  
 gcaagcaaga cctcccttcac cggccccatc ctgctgctg gaaccgtctc ctccatcaac 1740  
 ctcagccagc tcgaccccaat tgcagtgtat gaggtgaagc tgctgcctca caaccagcat 1800  
 ggagatggca atgcacacagt cgcgtttgtg tctttgaggg gacatctga gaggacagggc 1860  
 atcgteatcg gcatccacat cggggtcact tgcatcatct tctgtgtctt cttectcttg 1920  
 ttcggccaaa gggcagggtt ctcctgtgt aaagatgtgg aaaaccagct gtcccccctcca 1980  
 cagggtcccc gggccagag ggaccctggc attctggccc taaatggggc gagacgggg 2040  
 cagcggggcc agctggggcc agacgagaaaa cgtgtggata tgaaggagct ggagcagctg 2100  
 ttcggccccc ccagcgcagc agggcagccg gacccagac ccacagatcc tgcagcccc 2160  
 gtcgggtgtt aggacccca gcttccttgc ctgcccacttc aggggtgggg cctgtatggag 2220  
 gggaaagacga cggaggcggaa gaccacagag gccacggctc cctgcggccgg cctggccgct 2280  
 gccccccaccc ccccgatgg agggccctggc ctccctcactg aaggccagggc ttccaggcc 2340  
 gcagcggggcc gggttaccca gccagctcac tggaaacagt agccagtgtc tggcaggctc 2400  
 cagagggtgg acggagcggg gcccattctc aggtcaaaag caagatttct actgtcatgt 2460  
 gggattttggaa tggtcctggg ggctccccag catttctatc ctgactgcctt cttgggttgg 2520  
 caaaaacccaa ggcagcccttgc acagggaccc cccggccctaa acacccatca ggagttggag 2580  
 cagttcctgc aggacccctgt tccctccctg ggctgacgccc ccctgcctc tgcctgggta 2640  
 cccacatgac ttggaaactga actaacatcc ttctttaaaaa agaaaaaactt aaaaaaaaaa 2700

&lt;210&gt; 8

&lt;211&gt; 2321

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7701502CB1

&lt;400&gt; 8

gtataagtat atcccatgca gcatttcgga catacttata caaaaaaaaaca taaaataaaaa 60  
 tactttatat tggtgtgaa agctttaacc aggcgtctg cattttaccc gacaagacca 120  
 gcccccttcca gtaacggtcc cttccctgtt acccaccaccc tccaccacca gtggcccaaga 180  
 taagggttagc tcactgtggaa cttctcaactt ctgctttcc gatgacccaggg gggccagagg 240  
 gagacccccc accccactac cccacccca gccaggcccc caggatagac acaagcctgc 300  
 ccttcctcc tacacaggga gctctggaa gaccagctca acaagcccaa ggccaggctg 360  
 ggcccccagg agacccagag gacaactggg caagggtggc cggagagtgtt gggggaaaggc 420  
 aaaggaggatc tggtagctca gctgtctgaa ctcatttcat gcatcaggccc ccagggtctca 480  
 gcttcggccct tcgggttccc cttctgccaa gagccctgag ccactcacag cacgaccaga 540  
 gaacaggccct gtcctaggca ggccctgcgc ctcctatgcg gagatgttac tggcactgt 600  
 gctgtcctcg ctgctggcg ggtcccgaggc tatggatggg agattctgga tacgagtgtca 660  
 ggagtcaatg atgggtccggg agggccctgtt catctctgtt ccctgctt tctccttaccc 720  
 ccgacaggac tggacagggtt ctacccaccc ttatggctac tgggtcaaaag cagtactga 780  
 gacaaccaag ggtgtctctg tggccacaaa ccaccaggat cgagagggtgg aatgagcac 840  
 ccggggccga ttccagctca ctggggatcc cgccaaagggg aactgctctt tgggtatcag 900

agacgcgcag atgcaggatg agtcacagta cttcttcgg gtggagagag gaagctatgt 960  
 gagatataat ttcatgaacg atgggttctt tctaaaagta acagccctga ctcagaagcc 1020  
 ttagtgcac atccccgaga ccctggagcc cgggcagccg gtgacggta tctgtgttt 1080  
 taactggcc tttgaggaat gtccacccccc ttcttcctcc tggacggggg ctgcctctc 1140  
 ctcacaaggaa accaaaccaa cgacccccc cttctcagtg ctcagcttca cgcggcagacc 1200  
 ccaggaccac aacaccgacc tcacctgca tggacttc tccagaaagg gtgtgagcgt 1260  
 acagaggacc gtccgactcc gtgtggcta tgccccaga gacccgtta tcagcatttc 1320  
 acgtgacaaac acgcccggcc tgagccca gccccaggg aatgtcccat accttggaaac 1380  
 ccaaaaaggc cagttcttcg ggtctcttg tgctgtcagc agccagcccc ctgcccacact 1440  
 gagctgggtc ctgcagaaca ggttccttc ctcgtcccat ccctggggcc ctagacccct 1500  
 ggggctggag ctgccccggg tgaaggctgg ggattcaggc cgttacaccc gcccggcgg 1560  
 gaacaggctt ggctccca ggcggccct ggacccctct gtgcagtata ctccagagaa 1620  
 cctgagatgt atggttttcc aagcaaaacag gacagtctg gaaaacctt ggaacggcac 1680  
 gtctctccca gtactggagg gccaaaggctt gtgcctggtc tgtgtcacac acagcagccc 1740  
 cccagccagg ctgagctgca cccagggggg acaggttctg agccctccca agccctcaga 1800  
 ccccggggtc ctggagctgc ctccgggttca agtggagcac gaaggagagt tcacctgcca 1860  
 cgctcgccac ccaactgggtt cccagcacgt ctctctcagc ctctccgtc actataagaa 1920  
 gggactcatac tcaacggcat tctccaaacgg agcggttctg ggaatcgca tcacggctct 1980  
 tcttttcctc tgccctggccc ttagatgtta ggttaagagg aggctggcg gggctgggg 2040  
 cctggaccctt ggaaggggg aggtgttgca ggcggaaaga gtgaaggctg tgatcaacgc 2100  
 agtataccctc tggaggttac atgagtaaac agcaaaactgt tctcataat gcagaatgtt 2160  
 gtccaaactga caaaactgcgt ctgttccca gaggaaatgc tgaggcagt cacggcccaa 2220  
 gcgaaagtgtt tcttggtaatt aggccacagct ggacccctt agtaataata tgaacctgtt 2280  
 atcaattaaa cagctgacca attgttaaaa aaaaaaaaaa a 2321

<210> 9  
 <211> 2112  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 5727379CB1

<400> 9  
 atgcttcaca cggccatatac atgctggcag ccattctgg gtctggctgt ggtgttaatc 60  
 ttcatgggat ccaccattgg ctgccccgct cgctgtgagt gctctccca gaacaaatct 120  
 gttagctgtc acagaaggcg attgatcgcc atcccagagg gcatcccat cgaaaaccaa 180  
 atcttggacc tcagtaaaaaa caggctaaaaa agcgtcaacc ctgaagaatt catatcatat 240  
 cctctgttgg aagagataga ctgtgttgc aacatcattt ccaatgttgg accaggagca 300  
 ttcaacaatc tctttaaacct gcttccctc cgcctaaaag gcaatcgctt aaagctggtc 360  
 cctttggggag tattcacggg gctgtccat ctcactaagg ttgacattag tgagaataag 420  
 attgtcattt tactagacta catgttccaa gatctacata accttgaagtc tctagaatgt 480  
 ggggacaatg atttggtttataatcacac agggcatca gtgggtttt tagcttggag 540  
 cagctcaccc tggagaaatg caacttaaca gcaatccaa cagaaggccct ctccaccc 600  
 cgcagccctca tcagccctgca tctgaagcat ctcaatatca acaatatgcc tggatgtatgcc 660  
 tttaaaagat tggccacccctt gaaacaccta gagattgact attggccctt actggatatg 720  
 atgcctgcac atagccctca cggtctcaac ctccatcccc tttcaagtcaac caacaccaat 780  
 ctgtctactg tacccttctt tgcctttaaa caccctggcgttccatataatgt gggggccca 840  
 cttcgcacca ttgagctca ctccctccaa gggctccgtt ctcaacccgtt gctcaatgtt 900  
 tctcagaacc tgctggaaac ttggagag aatgtcttctt ctcccttag ggctctggag 960  
 gtcttggcgtt ttaacaacaa ccctctggcc tggactgtcc gcttctctg gatcttgcag 1020  
 cgacagccca ccctgcgtt tggggccag caacccatgt gtcgtggccc agacaccatc 1080  
 cgtgagaggt tttcaaggaa ttccatagc actggccctt cttttactt tacctgcaaaa 1140  
 aaacccaaaaa tccgtaaaaa gaagttgcag catctgtcgtt tagatgaagg gcagacagtc 1200  
 cagcttagaaat gcaatgcaga tggagaccccg cagccctgtca tttccctgggt gacaccccg 1260

aggcgtttca tcaccaccaa gtccaatgga agagccacccg tggcaccttg 1320  
 gaaatccgc ttgcccaggag tcaagacagc gggatgtatg tttgcacgc tagcaatgc 1380  
 gctggaaatg atacccctcac accctccctt actgtgaaag gattcgcetcc agatcggttt 1440  
 ctttatgcga acaggacccc tatgtacatg accgactccca atgacaccat ttccaatggc 1500  
 accaatgcacca atacttttc cctggacctt aaaacaatac tggtgtctac agctatgggc 1560  
 tgcttcacat tcctggaggt ggttttattt tggggatc tccctttgt gtggagccga 1620  
 gggaaaggca agcacaaaaa cagcattgac cttgagtatg tgcccgagaa aaacaatggt 1680  
 gctgttgggg aaggggaggt agctggaccc aggaggttca acatgaaaat gatttgaagg 1740  
 cccaccccttc acattactgt ctctttgtca atgtggtaa tcagtaagac agtatggcac 1800  
 agtaaattac tagattaaga ggcagccatg tgcagctgcc cctgtatcaa aagcagggtc 1860  
 tatggaaagca ggaggacttc caatggagac tctccatcga aaggcaggca ggcaggcatg 1920  
 tgcagagcc cttcacacag tgggatacta agtggggcg ttgcaaataat tggcggtctg 1980  
 gggatcttag taatgaacct gaatatttgg ctcacactca cggacaatta ttcagcattt 2040  
 tctaccactg caaaaaaaaaaaa aaaaaaaaaat aaaaaaagaac tacctacagt gtaggattta 2100  
 catattaaaaa ag 2112

&lt;210&gt; 10

&lt;211&gt; 1419

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 5831801CB1

&lt;400&gt; 10

gtacgagttg gtcgagctcc atcactgata cggacgccag tggctggaa agcttacatg 60  
 taccctcctc ctcaggaaaga aggagtgtta ggtggagaca gggatagttt ggacacaaaag 120  
 tccagtttaag tgggtgtccc agctcagtgc tctgagccca gatccctcatc tccctgggta 180  
 gtgaggctca gcacagacaa gcaaccaact gctgggtgc cggtgcggcc catgttggaa 240  
 cctgagttgg agattatctc ctaagcagat acctgtctt caaactgggg atgttagggct 300  
 tggaaacttag aaaaatgccag gtcgaggaga gaggaaagaa caagtccagc aatacacaga 360  
 gctctgtgta ttcagagggg agttggcagg gttgtgttcg ggcagagaaaa ctccgagtgg 420  
 tacaaggggg acgtgcccag agtggagaaaa tcatgctaat tgcact agagctggag 480  
 aacgccaccc aaaaatgaaga gagaaggggg agccctgtcc agagctcca gggccctgcg 540  
 cttgtctct tttgtctacc ttctctgtat ccagacagac cccctggagg gggtaacat 600  
 caccagcccc gtgcgcctga tccatggcac cgtggggaaag tggctctgc tttctgtgca 660  
 gtacagcagt accagcagcg acaggcctgt agtgaagtgg cagctgaagc gggacaagcc 720  
 agtgaccgtg gtgcagtcca ttggcacaga ggtcatcgcc accctgcggc ctgactatcg 780  
 agaccgtatc cgactctttg aaaaatggctc cctgcttctc aegacactgc agtggcccg 840  
 tgagggcacc tatgaggtcg agatctccat caccgacgac accttactg gggagaagac 900  
 catcaacattt actgttagatg tggccatattt gaggccacag gtgttgggtgg cttcaaccac 960  
 tggctggag ctcagcgagg cttcacctt gaactgctca catgagaatg gcaccaagcc 1020  
 cagctacacc tggctgaagg atggcaagcc cttctcaat gactcgagaa tgcctctgtc 1080  
 ccccgacccaa aagggtctca ccatcaccccg cgtgcctatg gaggatgacg acctgtacag 1140  
 ctgcacgggtt gagaacccca tcaagccaggg cccgcacccg cctgtcaaga tcaccgtata 1200  
 cagaagaagc tcccttaca tcatcttgc tacaggaggg atcttctcc ttgtgacctt 1260  
 ggtgacagtc tggctctgtt gggaaaccctc caaaaggaaa cagaagaagc tagaaaagca 1320  
 aaactccctg gaatacatgg atcagaatga tgaccgcctg aaaccagcccc caaaggacca 1380  
 cagccccctag gctggacaac tccttcagtt tgcccgaccc 1419

&lt;210&gt; 11

&lt;211&gt; 6374

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7195090CB1

&lt;400&gt; 11

atggatgaag ctggaaaacca tcattcttag caaactaaca caaggacaaa aaaccaaaca 60  
ccgcatgttc tcactcatag gtacattatt ccatcttgc agaagctcg tgctgggttt 120  
taccgctgcg tggtcgaaa cagaatggga gcactcctgc aaagaaaatc agaagttcaa 180  
gtcgatata tggaaagttt catggatacg gaccagagga aaacagtttc tcaaggacgt 240  
gcagcgattc taaacctgtc gcccattcacc agctacccca gacctcaagt gacttgggtt 300  
agagaaggc acaagattat tccaagcaac agaatagcca tcacattgga gaatcagctg 360  
gtgatctcg ccaccacaac cagtgtatcc gggcataact acgtgcaggg cgtgaatgag 420  
aaaaatggag aaaacaagac aagccattc attcattga gcatacgaaag agatgttggc 480  
acacctgaaa ccatggcccc accattgtg gttcccccgg gcaacagaag tgggtggct 540  
ggatccagtg agaacacctt ggaatgtata gccagtggca ggctgtgg ggacctgagt 600  
gtgacctgga agagaatgg aagtggaaatc accagtggcc tccacagctt tggaaagacgc 660  
ctcaccatca gcaacccgac gtccgcggac accggccat acgtctgcga ggcggcgctg 720  
ccggggagcg ctgttgaacc ggcggggcg acggcccttc ttttcatcat agagccacca 780  
tattttactg ctgagcccgaa gatgtggatt tcaagctgaa tagaagaaac tggacatc 840  
ggatgtcaag ccattggggtt ccccttccc accctccatg ggtacaagga tgcacatctcc 900  
atcagcagggc tccagaatcc tcgatacaaa gtgctgcgc gcgaggcct ggcacatccag 960  
aagctgcgtc cagaggactc cggaaatcttc cagtgttgcg ccagcaatga aggaggggag 1020  
atccagaccc acacctaccc gatgtaaacc aatatcgctc cagtgttgc acagcggcca 1080  
gtggacacca cagttactga cgggatgaca gccattctaa ggtgtgagggt gtcggggct 1140  
cccaaacccg ccattcacctg gaaaagagaa aaccacattc tggccagtgg ctctgtccgg 1200  
attccttaggt tcatgtttct tgaatcgggg ggtctacaga tcgcggccgtt cttcatccag 1260  
gatgcggca actacacctg ctatgcggcc aacacagagg gctccctgaa tgcacatggcc 1320  
acgctcaactg tggaaatcg gacgtccatc gtccacccctc ctgaggacca cgtgggtatt 1380  
aaggggacca cggccacgct gactgtggt gcccacatg acccccccgggtt ttcactccgc 1440  
tacgtttgga agaaggacaa cgtggccctg actccatcg gacgtctcg gatgtgggt 1500  
gagaaggacg ggtcccttct catcagcccg acgtgttcg ggcacatcg tgactacagc 1560  
tgcgagattt ttctgttggg agggaaatgac tccaggatgg cccggctgaa agtggattgaa 1620  
ctgcctcatt cacctcagaa cctcctggtc agccctaatt tttccacag ccacgcccgt 1680  
gtgctctttt ggtccggcc ctgttggaa aacagtcttca ttcttttata catcgtggag 1740  
ctctctgaaa acaactctcc atggaaagggtt catctgtcaa acgttggccc tgagatgaca 1800  
ggcgtccacccg tgagtggcct gactccggct cgtacccatc aattccgggtt gtgcgggtg 1860  
aatgaagtgg gcagggccca gtacagcgcc gagacaagca ggttgcgtct acctgaagaa 1920  
ccacccatgt ctccccccgaa aatatacgat ggcagttggc ggactaatca gtccattatg 1980  
gtccagttggc agccacccccc agaaacagag cacaacgggg tggcgttgg atacatcctc 2040  
aggtaaccgc tggctggcct tcccgagag taccagccgc ggaacatcac cagcccccggag 2100  
gtgaactact gcctgggtac agacctgatc atctggacac agtgcgtatc acagggtggcg 2160  
gcgtacaacg gggccggctt gggcgttcc agcaggccgc tgaccgcgtt cacccatcg 2220  
ggagtgccttcc cccggcccccc gcaacatgtt cagacggaaag ccgtgaactc caccaccatt 2280  
cagttctgtt ggaacccctcc gcctcagcg tttatcaatg gcatcaacca gggataacaag 2340  
cttctggcat ggcggccaga tgccccccgg gctgtcaactg tggctactat tggcccttgc 2400  
ttccacggag tccaccatgg acacataacg aacctgaaaga agttaaccgc ctacttact 2460  
tccgttctgtt gcttcaccac ccctggggac gggcccttccca gcacacctca gctggcttgg 2520  
actcagggaaag acaaaccagg agctgtgggaa catctgatgtt tcacagagat tggacaca 2580  
tctctcaagg tcagctggca ggagccctgtt gagaatggatc gcatcattac tggctatcg 2640  
atctcttggg aagtgtacgg caggaacaac tctcgctca cgcacaccctt gaaacacgc 2700  
acgcacgatgtt acaagatcca aggccctctca tctctcaccat cctacaccat cgacgtggcc 2760  
gctgtactg cctgtggccac tggcctgggtt acttcatttccca ccatttcttc tggagtgccc 2820  
ccagaccccttc ctggggcccccc atccaaaccttgc tgcatttccca acatcagcccg tgcctccggcc 2880  
acccttcaatgtt tccggccagg ctatgcggg aaaacgttccca tctccagggtt gattgttgg 2940  
ggcgggggtgg gagctatcggtt ccacggggag ggtgggtca ccctctatgaa agaggagaat 3000  
gaggctgatgtt cccagatgtt gggatccca aacctcacac cctacactca ctacagattt 3060

cgaatgaac aagtgaacat tgttggccg agccctaca gtccgttcc cgggtcata 3120  
 cagaccctgc aggccccacc cgacgtggct ccaaccagcg tcacggccg tactgccagt 3180  
 gagaccagcc tgcggcttgc ctgggtgccc ctgcggatt ctcagtacaa cgggaacccc 3240  
 gagtcgtgg gtcacaggat taagtactgg cgctcagacc tccagtcctc agcagtggcc 3300  
 caagtgcgtca gtgaccggc ggagagagaa ttacccatcg aggagctgga ggagtggatg 3360  
 gaatacagac tgcagatgca ggccttcaac gccgtgggg ctggccgtg gagcgaggtg 3420  
 gtgcggggcc ggacgcggg gtcagttct tcagccgccc ctgagaacgt gtcagccgag 3480  
 gtcgtcagct cgaccaggat ttactgaca tggacatccg tgccggaaaca ggaccagaat 3540  
 gggctcatac tgggtacaa gatctgttc cggccaaag acctggatcc cgagccagg 3600  
 agccacatcg tgcgaggaa ccacacgcag tggccctgc tggcaggct ggcgaatgc 3660  
 gtgcgtctacg agctccaggat gtcggcttc accegcatcg ggaacggggt cccagcacg 3720  
 cccctcatcc tggagcgcac caaagacgat gccccaggcc caccagttag gtcgtgtt 3780  
 cccgaagtgta gactcaccc cggtggata gtgtggcaac ctccggagga gccccacggc 3840  
 atcatcctgg ggtaccaggat tgcctaccgc ctggccagca gcagccccca caccttcacc 3900  
 accgtggagg tcggccac agtgaggcag ttacacagca cgcacccggc cccggagtcc 3960  
 gcatacatct tcaggctgtc cgccaaagacg aggccaggct ggggggagcc actggagggc 4020  
 accgtcatca ccaccggagaa gagagagcgg cccgcacccc ccagagagct cctggtgccc 4080  
 caggcagaag tgaccgcacg cagccctccgg ctccagttggg tccggggcag cgacggggcc 4140  
 tccccatcc ggtacttcac catgcagggtc cgagagctgc ctgggggtga gtggcagacc 4200  
 tactctctgt ccatacggaca tgaggcaca gcatgcgtcg ttgacagact gaggcccttc 4260  
 acctcttaca agctgcgcct gaaagccacc aacgacattt gggacagtga ttgcgttca 4320  
 gagacagagg cggtagccac gtcggatgtt gttccaggag agccccccggg atctgtctca 4380  
 ggcacgcac acaccacgtc ctctgtctg atacagtggc agccctccgg ggcacggaaagc 4440  
 ctgaatggcc ttcttcaggg atacaggatc tactacaggg agctggatg tgaagccggg 4500  
 tcaggcactg aggccaaagac gtcaaaaaac cctatacgat tacatgcgtga gtcacagggc 4560  
 caaaggcagct tcaagacggt gaacagcgcg tccacatcg cgtgtgtga actaacacat 4620  
 taaaagaagt accggcgcta tgaagataa atgaccgcct ataacatcat cggcgagagc 4680  
 ccagccagcg cgccctgtgg ggtcttgc ggcggggctg cccggccat gggccggcag 4740  
 aacgtgcagg tgacccact cacggccacg cagctggagg tcatgtggga cccaccaccc 4800  
 cccggagggcc agaatggaa catccaaggc tacaagattt actactggg ggcacacgc 4860  
 cagaacgaaa cggagaaaat gaaggtcctc ttccctcccg agccctgtgt gaggctgaag 4920  
 aacctgacca gcccataccaa gtacctgtc agcatatcg cttcaacgc cgccggagat 4980  
 ggacctaaga gtgaccccca gcagggggcgc accccaccagg cccggccctgg gggcccccagc 5040  
 ttctggcgt tctcagaaat aaccccccacc acgctcaacg tgcgtgttggg cggcctgcg 5100  
 gcccggccac gcatctgtca gggctatcgg gtgggttacg agccctggc ccctgtacaa 5160  
 ggggtgagca aggtgggtac cttggaaatg agagggaact ggcacgcgtg gctgaagggt 5220  
 cggggacctca ccaagggagt gacattttc ttccgtgtcc aagcggggac catcacctac 5280  
 gggccggcgc tccaagccaa tatcacagcc gggccagccg agggatcccc gggctcgcc 5340  
 agagatgtcc tggtcaccaa tccggctct gaactgacgc tgcgtggac tgagggacac 5400  
 tctggcgaca cacctaccac gggtatgtg atcgaggccc ggcctcaga tgaaggctta 5460  
 tgggacatgt ttgtgaagga catcccgccg agccacat cctacaccct cagcctggat 5520  
 aacgtccggc aaggagtgc ttacgatgc cgggtgggtt ctgtgaatga ggcgggctac 5580  
 ggggagccca gcaacccctc cacggctgtc tcaatcaag tggaaagcccc attctacgag 5640  
 gagttgggtt tccctctgtt gatggctgtc tccagccctga tgcgtatccct gtcgggttg 5700  
 ttccggctcg tcctgcacgg gcaaaaaaaat aagtataaga actgcacccac agggaaagggg 5760  
 atctccacca tggaggatgc tgcgtgttgc gacaacggg gattgtgtc cctggagctc 5820  
 agcagccgccc acctcaatgt caagagcacc ttctccaaga agaacgggac caggtcccc 5880  
 cccggccata gccccggccg cttgcactac tcagacgagg acatctgaa caagtacaac 5940  
 ggcggccgtgc tgacccggag cttggccctc aaggagaatg cggcagatgc atcagaatct 6000  
 gaggccacgg actctgacta cgaggacgcg ctggccaaagc actccctgtt gacccactac 6060  
 atgagcgcacc ccacccatca caactcatgg aagcgcaggc cccaggccgg cgcacccgtc 6120  
 cccgcacatgt tagctatccct cctgcacggc aacccgtctg cgtacctaag tgcgtgttccc 6180  
 cgtggggccatc cgtccctgtt gcatggatcc agtctgaaatg gtgaggacaa cgtggaaact 6240  
 catgagctga gcttgcggc tggacacgt tccctcccg cgtcacccctc tggtttaggg 6300  
 agccgtcagg tccctaaacg ttccctacaa tttttctgtaa aattgtgcag aaaaacagat 6360  
 ctcattaaaga aaaa  
 6374

<210> 12  
<211> 1350  
<212> DNA  
<213> *Homo sapiens*

<220>  
<221> misc\_feature  
<223> Incyte ID No: 7473826CB1

atgacaagcg tgaccataaa aggctctgga tttgccgtt cttctgcagg tggaaaagtc 60  
 cttatgggtc atttccatg taaaaggta tcagtgaatt atacggccat tgaatgtgaa 120  
 acatccctg ctgccccaaa gcttggat gttagatctt taatacatgg agtgcctgcc 180  
 cagtgcagg gaaactgcac ctttcatac ttagaaagca tcactccta cataacagga 240  
 gtcttccaa actctgtcat aggtctgtt aaagttctt ttgaaggaga aggtttgggg 300  
 actgtttgg aggacatttt tggacatcat tctgttagt ttgtgggtgg aagtaaaagc 360  
 ttggctctgg gaaacctgac tgcgtcggc cccccagtag catctctatc accaacttct 420  
 ggaagcattt gttgttggaaac tacactgggt atcacagggaa atggcttcta tccaggcaac 480  
 actacagtc ctattggggaa tgaaccttgc caaaattttt ccatcaaccc caatgaagtc 540  
 tactgcccga ctcccgctgg gaccactggaa atggctgttgc taaaatctt tgtaatacaca 600  
 attgcttatac cacccttgc ttttacatat gcccgtggg atactccatt ttcagaggg 660  
 attatcccaa gcagagggtcc accaggaaact gaaatttggaa tcactggatc caactttggc 720  
 tttgagatct tggaaatctc cgtgtatgata aataacatc agtgaatgt aaccatggcc 780  
 aatgatagtg tgggtgcagtg catcggtggg gatcatgtt ggggcacatt tccgtttatg 840  
 atgcatcata agacaaaagg ctgcggccacg tccacagttg tatttggata cccgcttaat 900  
 attcaaaata ttaatccaag ccaaggtagt cataagtatg caagaacctg cagatatttt 960  
 gtttcagtaa acacttattt agaaaccttta ttttagaaaat aatatgttac tcaaacacat 1020  
 aaaatttcta gtgtttttt tttttttttt ttttagacag agtctcgctc 1080  
 ggtcgccca gctggagtgc agtggcgcga tcttggctca ttgcaagctc tgcctcccaag 1140  
 gttcacggca ttctcctgccc ttagccccc gagtagctgg gactacaggc gccccccacc 1200  
 atgcctggtc aattttttt gtttttttta gtagagacgg gtttccacca taaaattttt 1260  
 tagcacaata caaatacttg gttttatgtt gcaaaaat aagttgtaa gtgtataatt 1320  
 atttatttca aaatactaaa cttaaaata 1350